

1

SEARCH REQUEST FORM
Scientific and Technical Information Center

Access DB# 51683

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: Beverly C 4999

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: _____

Date Completed 09-28-01

Searcher Prep & Review Time _____

Clerical Prep Time _____

Online Time _____

Type of Search

NA Sequence (#) _____

AA Sequence (#) _____

Structure (#) _____

Bibliographic _____

Litigation _____

Fulltext _____

Patent Family _____

Other _____

Vendors and cost where applicable

STN _____

Dialog _____

Questel Orbit _____

Dr. Link _____

Lexis/Nexis _____

Sequence Systems _____

WWW/Internet _____

Other (specify) CGN

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2001, 10:23:44 : Search time 51.92 Seconds
(without alignments)
11.676 Million cell updates/sec

Title: US-09-551-645-1
Perfect score: 49
Sequence: 1 TFIGAATDF 10

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /cgnl_9/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /cgnl_9/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /cgnl_9/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /cgnl_9/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /cgnl_9/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /cgnl_9/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /cgnl_9/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /cgnl_9/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /cgnl_9/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /cgnl_9/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /cgnl_9/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /cgnl_9/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /cgnl_9/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /cgnl_9/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /cgnl_9/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /cgnl_9/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /cgnl_9/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /cgnl_9/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /cgnl_9/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /cgnl_9/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /cgnl_9/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /cgnl_9/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	98.0	1612	19	AAW65088
2	34	69.4	455	18	AAW17830
3	34	69.4	468	20	AAV34877
4	34	69.4	1267	11	AAW04232
5	33	67.3	147	21	AAW30361
6	33	67.3	162	21	AAW30360
7	33	67.3	179	21	AAW30359
8	33	67.3	248	22	AAW79437
9	33	67.3	531	19	AAW79067
10	32	65.3	143	21	AAW41069
11	32	65.3	253	20	AAW65370

12	32	65.3	253	20	AAW65371	Duroc alpha melano
13	32	65.3	253	20	AAW86366	Wild boar alpha me
14	32	65.3	253	20	AAW86367	Melshan alpha mela
15	32	65.3	253	20	AAW86369	Large White alpha-
16	32	65.3	317	14	AAW43571	Human MSH-R, Homo
17	32	65.3	317	15	AAW49725	Sequence of a poly
18	32	65.3	317	18	AAW19706	Melanocortin-1 rec
19	32	65.3	317	19	AAW79684	Melanocortin-1 rec
20	32	65.3	317	19	AAW37827	Human melanocyte S
21	32	65.3	317	19	AAW37821	Human melanocortin
22	32	65.3	317	20	AAW87865	Human melanocortin
23	32	65.3	317	20	AAW92439	Human MCI protein.
24	32	65.3	317	21	AAW18766	A human melanocort
25	32	65.3	317	21	AAW02001	Human melanocortin
26	32	65.3	317	21	AAW02002	Human melanocortin
27	32	65.3	317	21	AAW02003	Human melanocortin
28	32	65.3	317	21	AAW87420	Melanocortin-1 rec
29	32	65.3	382	21	AAW97019	Human melanocortin
30	32	65.3	382	21	AAW97020	Human melanocortin
31	32	65.3	382	21	AAW97021	Human melanocortin
32	32	65.3	382	21	AAW97022	Human melanocortin
33	32	65.3	382	21	AAW97023	Human melanocortin
34	32	65.3	382	21	AAW97024	Human melanocortin
35	32	65.3	382	21	AAW97025	Human melanocortin
36	32	65.3	382	21	AAW97026	Human melanocortin
37	32	65.3	382	21	AAW97027	Human melanocortin
38	32	65.3	382	21	AAW97028	Human melanocortin
39	32	65.3	620	21	AAW05950	Protein deduced fr
40	31	63.3	139	21	AAW55931	Arabidopsis thalia
41	31	63.3	139	21	AAW61199	Arabidopsis thalia
42	31	63.3	194	20	AAW37472	Amino acid sequenc
43	31	63.3	212	20	AAW82677	X. cumberlandia ty
44	31	63.3	212	20	AAW82678	X. cumberlandia ty
45	31	63.3	212	20	AAW82685	L. corniculatum ty

ALIGNMENTS

RESULT 1	
ID	AAW65088 standard; Protein; 1612 AA.
XX	
AC	AAW65088;
XX	
DT	28-SEP-1998 (first entry)
XX	
DE	R. prowazekii S-layer protein.
XX	
KW	Surface layer protein; S-layer; vaccine; typhus; spotted fever;
XX	
KW	Infection; diagnosis; disease.
XX	
OS	Rickettsia prowazekii.
XX	
PN	US5783441-A.
XX	
PD	21-JUL-1998.
XX	
PF	20-DEC-1993; 93US-0169927.
XX	
PR	20-DEC-1993; 93US-0169927.
XX	
PR	09-AUG-1991; 91US-0742128.
XX	
PA	(USNA) US SEC OF NAVY.
XX	
PI	Carl M, Ching W, Dasch GA, Dobson ME;
XX	
DR	WPI: 1998-427031/36.
XX	
DR	N-PSDB: AAV35235.
XX	
PT	Recombinant DNA encoding Rickettsia surface layer proteins - useful
PT	for diagnosing typhus and spotted fever and for preparing vaccines
PT	against them

XX Disclosure; Column 23-32; 20pp; English.
 PS
 XX
 CC This sequence represents the Surface layer (S-layer) protein from
 CC R. prowazekii strain Bren1. This sequence is useful for vaccination
 CC against typhus and spotted fever rickettsial infection or for diagnosing
 CC diseases caused by these bacteria. The surface layer protein antigens can
 CC be produced recombinantly in large quantities.
 XX
 SQ Sequence 1612 AA:
 QY 1 TFIGAATDT 10
 1111111111
 Db 1265 tfigaatacdt 1274

Query Match 98.0%; Score 48; DB 19; Length 1612;
 Best Local Similarity 90.0%; Pred. No. 0.24;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 2
 AAM17830 standard; Protein; 455 AA.
 ID AAM17830;
 AC AAM17830;
 XX
 DT 29-JUL-1997 (first entry)
 XX
 DE Thermophillic alkaline phosphatase.
 XX
 KM Alkaline phosphatase; thermophillic enzyme; label; assay.
 -XX
 OS Rhodothermus marinus strain ATCC 43812.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label/ Sig_peptide
 FT /note/ "signal peptide comprises amino acids
 1-20 when AP is expressed in E. coli"
 FT Protein 22..455
 /label/ Mat-protein
 FT /note/ "mature protein comprises amino acids
 21-455 when AP is expressed in E. coli"
 XX
 PN EP770678-A2.
 XX
 PD 02-MAY-1997.
 XX
 PF 24-OCT-1996; 96EP-0307692.
 XX
 PR 27-OCT-1995; 95US-0005965.
 XX
 PA (AMSH) AMERSHAM LIFE SCI INC.
 PI Davis M, Szasz J;
 XX
 DR WPI: 1997-238139/22.
 DR N-PSDB: AAT66463.
 XX
 PT Thermostable Rhodothermus marinus alkaline phosphatase - useful as
 PT enzyme label in immunoassays and nucleic acid assays
 XX
 PS Example; Fig 9; 24pp; English.
 XX
 CC The alkaline phosphatase (AP) (AAM17830) of Rhodothermus marinus
 CC (Rma) has a high pH optimum (10.8) and is thermostable, retaining
 CC 60% of its activity after 24 hr at 65 deg. It is also tolerant
 CC of other denaturing conditions, such as overnight incubation in 6 M
 CC urea at 65 deg. Recombinant AP can be produced in transformed
 CC host cells utilizing isolated nucleotide sequences (see also
 CC AAT66461-63). The properties of the enzyme make it suitable for
 CC use in numerous non-isotopic methods for the detection of proteins

CC and nucleic acids.
 XX
 SQ Sequence 455 AA:
 QY 1 TFIGAATDT 10
 1111111111
 Db 92 tyngaatacdt 101

Query Match 69.4%; Score 34; DB 18; Length 455;
 Best Local Similarity 70.0%; Pred. No. 48;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RESULT 3
 AAY34877 standard; Protein; 468 AA.
 ID AAY34877
 AC AAY34877;
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Amino acid sequence of a Chlamydia pneumoniae protein.
 XX
 KM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KM vaccine; neutralising epitope.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN W09927105-A2.
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98MO-1B01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 XX
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffiths R;
 XX
 DR WPI: 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 PT
 PS Page 819-820; Disclosure; 1912pp; English.
 XX
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 CC
 XX
 SQ Sequence 468 AA:
 QY 2 TFIGAATDT 10
 1111111111
 Db 233 figaatacdt 241

Query Match 69.4%; Score 34; DB 20; Length 468;
 Best Local Similarity 88.9%; Pred. No. 49;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 4
AAR04232
ID AAR04232 standard: protein; 1267 AA.
XX
AC AAR04232;
XX
DT 14-SEP-1990 (first entry)
XX
DE Rickettsia rickettsii p120 and 27.6kD surface proteins.
XX
KW Rocky Mountain Spotted Fever; vaccine; diagnosis.
XX
OS Rickettsia rickettsii R strain.
XX
FH Key Location/Qualifiers
FT Protein 1..595
FT /label=120 kD surface protein
FT Protein 596..1267
FT /label=27.631 kD surface protein
XX
PM US7429936-A.
XX
PD 13-MAR-1990.
XX
PF 11-NOV-1989; 89US-0135039.
XX
PR 11-NOV-1989; 89US-0429936.
XX
XX (USSH) NAT INST OF HEALTH.
PA Gilmore RD, Joste N, McDonald GA;
PI WPI: 1990-139718/18.
DR
XX Rickettsia rickettsii surface protein gene - used for diagnosis and for
PT producing protein for vaccines against Rocky Mountain Spotted Fever.
XX
PS Disclosure: ; Opp: English.
XX
XX Main disclosure is the DNA fragment encoding the 120kD surface protein
CC and the 756bp ORF encoding the 27.6 kD protein, however the specification
CC best available copy is of too poor a quality to produce a beneficial
CC index file.
CC Sequencing of the p120 gene showed an ORF starting with ATG at position
CC 226. The ORF has 3900 bp which encode 1300 amino acids resulting in a
CC protein with deduced mol.wt. of 132.611 kD. The amino acid sequence has
CC 31 potential N-glycosylation sites. The other ORF of 726 bp was found
CC about 100 bp downstream and on the opposite strand of the p120 gene.
CC Between the two ORF's is the 106 bp intergenic region which presumably
CC serves as a transcriptional termination site.
CC The recombinant protein can be used to vaccinate humans against RMSF. The
CC DNA can be used as a hybridisation probe for diagnosis of RMSF.
XX
SQ Sequence 1267 AA;

Query Match 69.4%; Score 34; DB 11; Length 1267;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGATATD 9
DB 915 tvgatvld 922

RESULT 5
AAG30361
ID AAG30361 standard: Protein; 147 AA.
XX
AC AAG30361;
XX
DT 17-OCT-2000 (first entry)
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 36283.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140605.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154019.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 67.3%; Score 33; DB 21; Length 147;
Best Local Similarity 60.0%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TPTGAIATDT 10
| | | | : | |
Db 129 tltgavskdt 138

RESULT 6

ID AAG30360 standard; Protein: 162 AA.

AC AAG30360;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 36282.

DE Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.
OS

```

XX      EP1033405-A2.
PN
XX
PD      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-0301439.
XX
XX      25-FEB-1999; 990S-0121825.
PR      05-MAR-1999; 990S-0123180.
PR      09-MAR-1999; 990S-0123548.
PR      23-MAR-1999; 990S-0125788.
PR      25-MAR-1999; 990S-0126264.
PR      29-MAR-1999; 990S-0126785.
PR      01-APR-1999; 990S-0127462.
PR      06-APR-1999; 990S-0128234.
PR      08-APR-1999; 990S-0128714.
PR      16-APR-1999; 990S-0129845.
PR      19-APR-1999; 990S-0130077.
PR      21-APR-1999; 990S-0130449.
PR      23-APR-1999; 990S-0130510.
PR      23-APR-1999; 990S-0130891.
PR      28-APR-1999; 990S-0131449.
PR      30-APR-1999; 990S-0132048.
PR      30-APR-1999; 990S-0132407.
PR      04-MAY-1999; 990S-0132484.
PR      05-MAY-1999; 990S-0132485.
PR      06-MAY-1999; 990S-0132486.
PR      06-MAY-1999; 990S-0132487.
PR      07-MAY-1999; 990S-0132863.
PR      11-MAY-1999; 990S-0134256.
PR      14-MAY-1999; 990S-0134218.
PR      14-MAY-1999; 990S-0134219.
PR      14-MAY-1999; 990S-0134221.
PR      14-MAY-1999; 990S-0134370.
PR      18-MAY-1999; 990S-0134768.
PR      19-MAY-1999; 990S-0134941.
PR      20-MAY-1999; 990S-0135124.
PR      21-MAY-1999; 990S-0135353.
PR      24-MAY-1999; 990S-0135629.
PR      25-MAY-1999; 990S-0136021.
PR      27-MAY-1999; 990S-0136392.
PR      28-MAY-1999; 990S-0136782.
PR      01-JUN-1999; 990S-0137222.
PR      03-JUN-1999; 990S-0137528.
PR      04-JUN-1999; 990S-0137502.
PR      07-JUN-1999; 990S-0137724.
PR      08-JUN-1999; 990S-0138094.
PR      10-JUN-1999; 990S-0138540.
PR      10-JUN-1999; 990S-0138847.
PR      14-JUN-1999; 990S-0138119.
PR      16-JUN-1999; 990S-0139452.
PR      16-JUN-1999; 990S-0139453.
PR      17-JUN-1999; 990S-0139492.
PR      18-JUN-1999; 990S-0139454.
PR      18-JUN-1999; 990S-0139455.
PR      18-JUN-1999; 990S-0138456.
PR      18-JUN-1999; 990S-0139457.
PR      18-JUN-1999; 990S-0139458.
PR      18-JUN-1999; 990S-0139459.
PR      18-JUN-1999; 990S-0139460.
PR      18-JUN-1999; 990S-0139461.
PR      18-JUN-1999; 990S-0139462.
PR      18-JUN-1999; 990S-0139463.
PR      18-JUN-1999; 990S-0139750.
PR      18-JUN-1999; 990S-0139763.
PR      21-JUN-1999; 990S-0139817.
PR      22-JUN-1999; 990S-0139899.
PR      23-JUN-1999; 990S-0140353.
PR      23-JUN-1999; 990S-0140354.
PR      24-JUN-1999; 990S-0140695.
PR      28-JUN-1999; 990S-0140823.
PR      29-JUN-1999; 990S-0140991.
PR      30-JUN-1999; 990S-0141287.
PR      01-JUL-1999; 990S-0141842.
PR      01-JUL-1999; 990S-0142154.
PR      02-JUL-1999; 990S-0142055.
PR      06-JUL-1999; 990S-0142390.
PR      08-JUL-1999; 990S-0142803.
PR      09-JUL-1999; 990S-0142920.
PR      12-JUL-1999; 990S-0142977.
PR      13-JUL-1999; 990S-0143542.
PR      14-JUL-1999; 990S-0143624.
PR      15-JUL-1999; 990S-0144005.
PR      16-JUL-1999; 990S-0144085.
PR      16-JUL-1999; 990S-0144086.
PR      19-JUL-1999; 990S-0144325.
PR      19-JUL-1999; 990S-0144331.
PR      19-JUL-1999; 990S-0144332.
PR      19-JUL-1999; 990S-0144333.
PR      19-JUL-1999; 990S-0144334.
PR      19-JUL-1999; 990S-0144335.
PR      20-JUL-1999; 990S-0144352.
PR      20-JUL-1999; 990S-0144632.
PR      20-JUL-1999; 990S-0144884.
PR      21-JUL-1999; 990S-0144814.
PR      21-JUL-1999; 990S-0145086.
PR      21-JUL-1999; 990S-0145088.
PR      22-JUL-1999; 990S-0145085.
PR      22-JUL-1999; 990S-0145087.
PR      22-JUL-1999; 990S-0145089.
PR      22-JUL-1999; 990S-0145192.
PR      23-JUL-1999; 990S-0145145.
PR      23-JUL-1999; 990S-0145218.
PR      23-JUL-1999; 990S-0145224.
PR      26-JUL-1999; 990S-0145276.
PR      27-JUL-1999; 990S-0145913.
PR      27-JUL-1999; 990S-0145918.
PR      27-JUL-1999; 990S-0145919.
PR      28-JUL-1999; 990S-0145951.
PR      02-AUG-1999; 990S-0146386.
PR      02-AUG-1999; 990S-0146388.
PR      03-AUG-1999; 990S-0146389.
PR      03-AUG-1999; 990S-0147038.
PR      04-AUG-1999; 990S-0147204.
PR      04-AUG-1999; 990S-0147302.
PR      05-AUG-1999; 990S-0147192.
PR      05-AUG-1999; 990S-0147260.
PR      06-AUG-1999; 990S-0147303.
PR      06-AUG-1999; 990S-0147416.
PR      09-AUG-1999; 990S-0147493.
PR      09-AUG-1999; 990S-0147935.
PR      10-AUG-1999; 990S-0148171.
PR      11-AUG-1999; 990S-0148319.
PR      12-AUG-1999; 990S-0148341.
PR      13-AUG-1999; 990S-0148565.
PR      13-AUG-1999; 990S-0148684.
PR      16-AUG-1999; 990S-0149368.
PR      17-AUG-1999; 990S-0149175.
PR      18-AUG-1999; 990S-0149426.
PR      20-AUG-1999; 990S-0149722.
PR      20-AUG-1999; 990S-0149723.
PR      20-AUG-1999; 990S-0149929.
PR      23-AUG-1999; 990S-0149902.
PR      23-AUG-1999; 990S-0149930.
PR      25-AUG-1999; 990S-0150566.
PR      26-AUG-1999; 990S-0150884.
PR      27-AUG-1999; 990S-0151065.
PR      27-AUG-1999; 990S-0151066.
PR      27-AUG-1999; 990S-0151080.
PR      30-AUG-1999; 990S-0151303.
PR      31-AUG-1999; 990S-0151338.
PR      01-SEP-1999; 990S-0151330.
PR      07-SEP-1999; 990S-0152363.
PR      10-SEP-1999; 990S-0153070.
PR      13-SEP-1999; 990S-0153758.
PR      15-SEP-1999; 990S-0154018.

```

PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158222.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159638.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 67.3%; Score 33; DB 21; Length 162;
Best Local Similarity 60.0%; Pred. NO. 24;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 TFIGAIATDT 10
DB 144 tligavskdt 153

RESULT 7
AAG30359
ID AAG30359 standard; Protein: 179 AA.
XX AAG30359;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SNO ID NO: 36281.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX 06-SEP-2000.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PF X?

PR 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137529.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147039.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 67.3%; Score 33; DB 21; Length 179;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TFGAATPDT 10
Db 161 tllgavskdt 170

RESULT 8
AAB79437
ID AAB79437 standard; Protein; 248 AA.
XX
AC AAB79437;
XX
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:390.
XX
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS Corynebacterium glutamicum.
XX
XX
PN MO200100844-A2.
XX
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000MO-IB00943.
XX
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.

PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031431.
 PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033005.
 PR 27-AUG-1999; 99DE-1040755.
 PR 31-AUG-1999; 99DE-1041572.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.
 XX (BADI) BASF AG.
 XX
 PA Pompejus M, Kroegeer B, Schroeder H, Zelder O, Habermayer G;
 PI
 DR MPI: 2001-061975/07.
 XX N-PSDB; AAF71554.
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 PS Claim 20; Page 701-702; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteolignoc
 CC or nonproteolignoc amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 CC
 XX
 SQ Sequence 248 AA;

Query Match 67.3%; Score 33; DB 22; Length 248;
 Best Local Similarity 66.7%; Pred. NO. 39;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 TFIGAATD 9
 I I I I I I I
 DB 65 tmlgavgtcd 73

RESULT 9
 AAW79067
 ID AAW79067 standard; Protein: 531 AA.
 XX
 AC AAW79067;
 XX
 DT 14-DEC-1998 (first entry)
 XX
 DE Sorghum bicolor (L.) Moench cytochrome P450ox monooxygenase.
 XX
 KW Cytochrome P450 monooxygenase; P450ox; Sorghum bicolor (L.) Moench;
 KW Sialapls alba; biosynthetic conversion; aldoxime; nitrile; cyanohydrin;
 KW cyanogenic glycoside; transgenic plant; resistance.
 XX
 OS Sorghum bicolor.
 XX
 PN WC0840470-A2.
 XX
 PD 17-SEP-1998.
 XX
 PE 05-MAR-1998; 98WO-EP01253.
 XX
 PR 08-DEC-1997; 97EP-0810954.
 PR 07-MAR-1997; 97EP-0810132.
 XX
 PA (NOVS) NOVARTIS AG.
 PA (UYRO-) UNIV ROYAL VETERINARY & AGRIC.
 XX
 PI Bak S, Halkier BA, Kahn RA, Moeller BL;
 XX
 DR MPI: 1998-520808/44.
 DR N-PSDB; AAV57472.
 XX
 XX
 PT Cytochrome P450 monooxygenase of the cyanogenic glycoside pathway -
 PT useful for the production of plants with improved nutritive value or
 PT pest resistance
 XX
 PS Example 6; Page 41-43; 32pp; English.
 XX
 CC The present sequence represents a cytochrome P450 monooxygenase from
 CC Sorghum bicolor (L.) Moench, designated P450ox. Cytochrome P450
 CC monooxygenase catalyses: (i) the conversion of aldoxime to a nitrile;
 CC and (ii) the nitrile to the corresponding cyanohydrin. DNA encoding
 CC cytochrome P450 monooxygenase can be used to obtain transgenic plants,
 CC for the purpose of improving the nutritive value or pest resistance of
 CC the plant. Cytochrome P450 monooxygenase catalyses the conversion of
 CC aldoximes to nitriles to cyanohydrins, which are the precursors of toxic
 CC cyanogenic glycosides, so staple food, such as cassava and lima beans,
 CC as well as animal feed such as white clover, can be rendered less toxic
 CC by blocking the cytochrome P450 monooxygenase activity. Introducing the
 CC enzyme to plants or to certain tissues could help reduce crop damage
 CC since the product is also toxic to insects, acarids and nematodes.
 CC
 XX
 SQ Sequence 531 AA;

Query Match 67.3%; Score 33; DB 19; Length 531;
 Best Local Similarity 70.0%; Pred. NO. 91;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 TFIGAATDT 10
 I I I I I I I I
 DB 325 tfigaldtss 334

RESULT 10
 AAB41069
 ID AAB41069 standard; Protein: 143 AA.
 XX
 AC AAB41069;
 XX
 DT 08-FEB-2001 (first entry)
 XX

DE Human ORFX ORF833 polypeptide sequence SEQ ID NO:1666.
 XX
 KM Human; open reading frame; ORFX; detection; cytosstatic; hepatotropic;
 KM vulnerability; antiparotatic; antiparkinsonian; neurotropic; neuroprotective;
 KM anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antineoplastic; antihypertensive;
 KM antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KM antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antineoplastic disease; coagulation;
 KM thrombosis; contraceptive.
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 PN
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 PI Shimketa RA, Leach M;
 DR WPI; 2000-602362/57.
 DR N-PSDB; AAC75278.
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 11: Page 1328; 5507pp; English.
 PS
 XX AAC74446 to AAC77606 encode the proteins given in AABA0237 to AABA43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosstatic; hepatotropic; vulnerability;
 CC antiparotatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antihypertensive;
 CC antihypertensive; antineoplastic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease, to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX Sequence 143 AA:
 SQ

Query Match 65.3%; Score 32; DB 21; Length 143;
 Best Local Similarity 75.0%; Pred. No. 34;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 FIGAINTD 9

DB 57 flgaatd 64
 1:111111
 ID AAM86370 standard; Protein; 253 AA.
 XX
 AC AAM86370;
 XX
 DT 26-MAR-1999 (first entry)
 XX
 DE Hampshire alpha melanocyte-stimulating hormone receptor partial protein.
 XX
 KM Porcine; wild boar; meishan; pietrain; large white; hamshire; duroc;
 KM differentiation; breed origin; alpha-MSHR; coat colour; stock purity;
 KM alpha melanocyte-stimulating hormone receptor; KIT.
 XX
 XX
 OS Sus scrofa.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1
 FT /note="unspecified"
 XX
 PN WO9854360-A1.
 XX
 PD 03-DEC-1998.
 XX
 PF 27-MAY-1998; 98MO-GB01531.
 XX
 PR 31-JAN-1998; 98GB-0001990.
 PR 30-MAY-1997; 97GB-0011214.
 XX
 XX (PIG-) PIG IMPROVEMENT CO UK LTD.
 PA
 PI Andersson L, Evans GJ, Gluftra E, Kljes J, Plastow GS;
 PI Wales R;
 DR WPI; 1999-070222/06.
 DR N-PSDB; AAV80685.
 XX
 XX Differentiating products from different animal breeds - by the
 PT analysis of alleles of breed-determinant genes, at the nucleic acid
 PT or protein level
 XX
 XX Claim 12: Fig 1; 101pp; English.
 PS
 XX A method has been developed for: (a) differentiating animals and animal
 CC products according to breed origin; (b) determining or testing the breed
 CC origin of a product; or (c) validating an animal product. The method
 CC comprises analysing a sample of the product for the allele(s) of at
 CC least one breed-determinant (BD) gene. The present invention also
 CC describes: (1) methods for determining the coat colour genotype of a pig
 CC by determining: (1) the allele(s) of the alpha melanocyte-stimulating
 CC hormone receptor (alpha-MSHR) gene; (1i) the amino acid sequence of an
 CC alpha-MSHR protein at positions associated with coat colour, or the size
 CC of the protein; (1ii) detecting which microsatellites (or other linked
 CC marker alleles), linked to the alpha-MSHR gene, or particular alleles of
 CC it, are present; and (1iv) analysing nucleic acid to determine if the KIT
 CC gene carries a polymorphism associated with the Belt genotype. The
 CC main method of the invention is applied to samples from fish, birds and
 CC mammals, especially pigs. Particular applications are confirming stated
 CC origin of meats; in quality control; for maintaining stock purity, and
 CC in breeding programmes (to confirm particular crosses). The method
 CC requires only very small samples and many samples can be screened
 CC quickly and inexpensively. The process can be made quantitative. The
 CC present sequence represents a partial protein sequence of alpha-MSHR
 CC from a Hampshire breed pig.
 CC N.B. The present sequence is said to be encoded by AAV80685, but does
 CC not appear to do so.
 CC
 XX Sequence 253 AA:
 SQ

Query Match 65.3%; Score 32; DB 20; Length 253;
 Best Local Similarity 75.0%; Pred. No. 65;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGAATD 9
 1:|||||
 Db 95 flgalavd 102

RESULT 12

AAW86371
 ID AAW86371 standard; Protein: 253 AA.

AAW86371:

DT 26-MAR-1999 (first entry)

XX Duroc alpha melanocyte-stimulating hormone receptor partial protein.

XX Porcine; wild boar; meishan; pietrain; large white; hampshire; duroc;
 KM differentiation; breed origin; alpha-MSHR; coat colour; stock purity;
 KW alpha melanocyte-stimulating hormone receptor; KIT.

XX alpha melanocyte-stimulating hormone receptor; KIT.

XX Sus scrofa.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "unspecified"

XX W09854360-A1.

XX 03-DEC-1998.

XX 27-MAY-1998; 98WO-GB01531.

XX 31-JAN-1998; 98GB-0001990.

XX 30-MAY-1997; 97GB-0011214.

XX (PIGI-) PIG IMPROVEMENT CO UK LTD.

XX Andersson L, Evans GJ, Giuffra E, Kijas J, Plastow GS;
 PI Wales R;

XX WPI: 1999-070222/06.

XX N-PSDB; AAV80686.

XX Differentiating products from different animal breeds - by the
 PT analysis of alleles of breed-determinant genes, at the nucleic acid
 or protein level

XX Claim 12; Fig 1; 101pp; English.

XX A method has been developed for: (a) differentiating animals and animal
 CC products according to breed origin; (b) determining or testing the breed
 CC origin of a product; or (c) validating an animal product. The method
 CC comprises analysing a sample of the product for the allele(s) of at
 CC least one breed-determinant (BD) gene. The present invention also
 CC describes: (i) methods for determining the coat colour genotype of a pig
 CC by determining: (1) the allele(s) of the alpha melanocyte-stimulating
 CC hormone receptor (alpha-MSHR) gene; (1i) the amino acid sequence of an
 CC alpha-MSHR protein at positions associated with coat colour, or the size
 CC of the protein; (1ii) detecting which microsatellites (or other linked
 CC marker alleles), linked to the alpha-MSHR gene, or particular alleles of
 CC it, are present; and (1v) analysing nucleic acid to determine if the KIT
 CC gene carries a polymorphism associated with the Belt genotype. The
 CC main method of the invention is applied to samples from fish, birds and
 CC mammals, especially pigs. Particular applications are confirming stated
 CC origin of meats; in quality control; for maintaining stock purity, and
 CC in breeding programmes (to confirm particular crosses). The method
 CC requires only very small samples and many samples can be screened
 CC quickly and inexpensively. The process can be made quantitative. The
 CC present sequence represents a partial protein sequence of alpha-MSHR

CC from a Duroc breed pig.
 CC N.B. The present sequence is said to be encoded by AAV80686, but does
 CC not appear to do so.
 XX Sequence 253 AA;

Query Match 65.3%; Score 32; DB 20; Length 253;
 Best Local Similarity 75.0%; Pred. No. 65;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGAATD 9
 1:|||||
 Db 95 flgalavd 102

RESULT 13

AAW86366
 ID AAW86366 standard; Protein: 253 AA.

AAW86366:

DT 26-MAR-1999 (first entry)

XX Wild boar alpha melanocyte-stimulating hormone receptor partial protein.

XX Porcine; wild boar; meishan; pietrain; large white; hampshire; duroc;
 KM differentiation; breed origin; alpha-MSHR; coat colour; stock purity;
 KW alpha melanocyte-stimulating hormone receptor; KIT.

XX alpha melanocyte-stimulating hormone receptor; KIT.

XX Sus scrofa.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "unspecified"

XX W09854360-A1.

XX 03-DEC-1998.

XX 27-MAY-1998; 98WO-GB01531.

XX 31-JAN-1998; 98GB-0001990.

XX 30-MAY-1997; 97GB-0011214.

XX (PIGI-) PIG IMPROVEMENT CO UK LTD.

XX Andersson L, Evans GJ, Giuffra E, Kijas J, Plastow GS;
 PI Wales R;

XX WPI: 1999-070222/06.

XX N-PSDB; AAV80681.

XX Differentiating products from different animal breeds - by the
 PT analysis of alleles of breed-determinant genes, at the nucleic acid
 or protein level

XX Claim 12; Fig 1; 101pp; English.

XX A method has been developed for: (a) differentiating animals and animal
 CC products according to breed origin; (b) determining or testing the breed
 CC origin of a product; or (c) validating an animal product. The method
 CC comprises analysing a sample of the product for the allele(s) of at
 CC least one breed-determinant (BD) gene. The present invention also
 CC describes: (i) methods for determining the coat colour genotype of a pig
 CC by determining: (1) the allele(s) of the alpha melanocyte-stimulating
 CC hormone receptor (alpha-MSHR) gene; (1i) the amino acid sequence of an
 CC alpha-MSHR protein at positions associated with coat colour, or the size
 CC of the protein; (1ii) detecting which microsatellites (or other linked
 CC marker alleles), linked to the alpha-MSHR gene, or particular alleles of
 CC it, are present; and (1v) analysing nucleic acid to determine if the KIT
 CC gene carries a polymorphism associated with the Belt genotype. The
 CC main method of the invention is applied to samples from fish, birds and

mammals, especially pigs. Particular applications are confirming stated origin of meats; in quality control; for maintaining stock purity, and in breeding programmes (to confirm particular crosses). The method requires only very small samples and many samples can be screened quickly and inexpensively. The process can be made quantitative. The present sequence represents a partial protein sequence of alpha-MSR from wild boar. The present sequence is said to be encoded by AAV80681, but does not appear to do so.

Sequence 253 AA:

Query Match 65.3%; Score 32; DB 20; Length 253;
 Best Local Similarity 75.0%; Pred. No. 65;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

2 FIGAINTD 9
 1:|||||
 Db 95 figalavd 102

RESULT 14
 AAW86367
 ID AAW86367 standard; Protein; 253 AA.
 AC AAW86367;
 XX
 XX
 DT 26-MAR-1999 (first entry)
 DE Melshan alpha melanocyte-stimulating hormone receptor partial protein.
 XX
 XX Porcine; wild boar; melshan; pietrain; large white; hampshire; duroc;
 KW differentiation; breed origin; alpha-MSHR; coat colour; stock purity;
 KM alpha melanocyte-stimulating hormone receptor; KIT.
 XX
 OS Sus scrofa.

Key Location/Qualifiers
 FH Misc-difference 1 /note="unspecified"
 FT
 FT
 XX
 XX
 PN WO9854360-A1.
 XX
 XX 03-DEC-1998.
 PD
 XX
 XX 27-MAY-1998; 98WO-GB01531.
 PF
 XX
 PR 31-JAN-1998; 98GB-0001990.
 PR 30-MAY-1997; 97GB-0011214.
 XX
 XX (PIGI-) PIG IMPROVEMENT CO UK LTD.
 PA
 XX Andersson L, Evans GJ, Gluffra E, Kljas J, Plastow GS;
 PI Wales R;
 DR WPI: 1999-070222/06.
 DR N-PSDB: AAV80682.
 XX
 PT Differentiating products from different animal breeds - by the
 PT analysis of alleles of breed-determinant genes, at the nucleic acid
 PT or protein level
 XX
 XX Claim 12: Fig 1; 101pp; English.

A method has been developed for: (a) differentiating animals and animal products according to breed origin; (b) determining or testing the breed origin of a product; or (c) validating an animal product. The method comprises analysing a sample of the product for the allele(s) of at least one breed-determinant (BD) gene. The present invention also describes: (1) methods for determining the coat colour genotype of a pig by determining: (1) the allele(s) of the alpha melanocyte-stimulating hormone receptor (alpha-MSHR) gene; (11) the amino acid sequence of an

alpha-MSHR protein at positions associated with coat colour, or the size of the protein; (111) detecting which microsatellites (or other linked CC marker alleles), linked to the alpha-MSHR gene, or particular alleles of CC it, are present; and (1iv) analysing nucleic acid to determine if the KIT CC gene carries a polymorphism associated with the Belt genotype. The CC main method of the invention is applied to samples from fish, birds and CC mammals, especially pigs. Particular applications are confirming stated CC origin of meats; in quality control; for maintaining stock purity, and CC in breeding programmes (to confirm particular crosses). The method CC requires only very small samples and many samples can be screened CC quickly and inexpensively. The process can be made quantitative. The CC present sequence represents a partial protein sequence of alpha-MSR from a Melshan breed pig. The present sequence is said to be encoded by AAV80682, but does not appear to do so.

Sequence 253 AA:

Query Match 65.3%; Score 32; DB 20; Length 253;
 Best Local Similarity 75.0%; Pred. No. 65;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

2 FIGAINTD 9
 1:|||||
 Db 95 figalavd 102

RESULT 15
 AAW86369
 ID AAW86369 standard; Protein; 253 AA.
 AC AAW86369;
 XX
 XX
 DT 26-MAR-1999 (first entry)
 DE Large white alpha-MSHR partial protein.
 XX
 XX Porcine; wild boar; melshan; pietrain; large white; hampshire; duroc;
 KW differentiation; breed origin; alpha-MSHR; coat colour; stock purity;
 KM alpha melanocyte-stimulating hormone receptor; KIT.
 XX
 OS Sus scrofa.

Key Location/Qualifiers
 FH Misc-difference 1 /note="unspecified"
 FT
 FT
 XX
 XX
 PN WO9854360-A1.
 XX
 XX 03-DEC-1998.
 PD
 XX
 XX 27-MAY-1998; 98WO-GB01531.
 PF
 XX
 PR 31-JAN-1998; 98GB-0001990.
 PR 30-MAY-1997; 97GB-0011214.
 XX
 XX (PIGI-) PIG IMPROVEMENT CO UK LTD.
 PA
 XX Andersson L, Evans GJ, Gluffra E, Kljas J, Plastow GS;
 PI Wales R;
 DR WPI: 1999-070222/06.
 DR N-PSDB: AAV80684.
 XX
 PT Differentiating products from different animal breeds - by the
 PT analysis of alleles of breed-determinant genes, at the nucleic acid
 PT or protein level
 XX
 XX Claim 12: Fig 1; 101pp; English.

A method has been developed for: (a) differentiating animals and animal products according to breed origin; (b) determining or testing the breed

CC origin of a product; or (c) validating an animal product. The method
CC comprises analysing a sample of the product for the allele(s) of at
CC least one breed-determinant (BD) gene. The present invention also
CC describes: (i) methods for determining the coat colour genotype of a pig
CC by determining: (i) the allele(s) of the alpha melanocyte-stimulating
CC hormone receptor (alpha-MSHR) gene; (ii) the amino acid sequence of an
CC alpha-MSHR protein at positions associated with coat colour; or the size
CC of the protein; (iii) detecting which microsatellites (or other linked
CC marker alleles), linked to the alpha-MSHR gene, or particular alleles of
CC it, are present; and (iv) analysing nucleic acid to determine if the KIT
CC gene carries a polymorphism associated with the belt genotype. The
CC main method of the invention is applied to samples from fish, birds and
CC mammals, especially pigs. Particular applications are confirming stated
CC origin of meats; in quality control; for maintaining stock purity; and
CC in breeding programmes (to confirm particular crosses). The method
CC requires only very small samples and many samples can be screened
CC quickly and inexpensively. The process can be made quantitative. The
CC present sequence represents a partial protein sequence of alpha-MSHR
CC from a Large White breed pig.
CC N.B. The present sequence is said to be encoded by AAV80684, but does
CC not appear to do so.
CC
XX

SO Sequence 253 AA;

Query Match 65.3%; Score 32; DB 20; Length 253;

Best Local Similarity 75.0%; Pred. No. 65;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGATATD 9

Db 95 flgalavd 102

Search completed: September 28, 2001, 12:26:03
Job time: 7339 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2001, 10:56:36 : Search time 25.24 Seconds
(without alignments)
8.158 Million cell updates/sec

Title: US-09-551-645-1
Perfect score: 49
Sequence: 1 TFIGAATDTR 10

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgnl_7/ptodata/1/laa/5A.COMB.pep:*
2: /cgnl_7/ptodata/1/laa/5B.COMB.pep:*
3: /cgnl_7/ptodata/1/laa/6A.COMB.pep:*
4: /cgnl_7/ptodata/1/laa/6B.COMB.pep:*
5: /cgnl_7/ptodata/1/laa/PCRUS.COMB.pep:*
6: /cgnl_7/ptodata/1/laa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	98.0	1612	1	US-08-169-927-2
2	34	69.4	455	2	US-08-738-172-4
3	32	65.3	317	1	US-07-866-979-6
4	32	65.3	317	1	US-08-671-525B-2
5	32	65.3	317	1	US-08-672-109B-2
6	32	65.3	317	2	US-08-842-045-2
7	32	65.3	317	2	US-08-466-906B-6
8	32	65.3	317	2	US-08-842-238-2
9	32	65.3	317	3	US-08-780-749A-4
10	32	65.3	317	3	US-08-706-281A-6
11	32	65.3	317	3	US-08-629-335B-2
12	32	65.3	317	4	US-09-201-746-6
13	30.5	62.2	164	4	US-08-911-319A-1
14	30.5	62.2	164	3	US-09-352-619-1
15	30	61.2	292	2	US-08-973-461A-4
16	30	61.2	292	3	US-08-648-010-4
17	30	61.2	572	4	US-09-401-476-4
18	29	59.2	44	4	US-08-687-590-21
19	29	59.2	44	4	US-08-687-590-22
20	29	59.2	45	1	US-08-056-200-107
21	29	59.2	45	2	US-08-800-644-107
22	29	59.2	89	1	US-07-987-272A-10
23	29	59.2	120	1	US-07-979-630-1
24	29	59.2	120	1	US-08-440-049-1
25	29	59.2	120	1	US-08-441-519A-1
26	29	59.2	120	3	US-08-970-865-3
27	29	59.2	120	4	US-09-363-573-3

28	29	59.2	120	5	PCT-US93-11292-1	Sequence 1, Appl1
29	29	59.2	120	5	PCT-US95-06918-1	Sequence 1, Appl1
30	29	59.2	189	1	US-08-709-912-13	Sequence 13, Appl1
31	29	59.2	189	2	US-09-047-370-13	Sequence 2, Appl1
32	29	59.2	295	2	US-08-464-517-2	Sequence 2, Appl1
33	29	59.2	295	2	US-08-246-361A-2	Sequence 2, Appl1
34	29	59.2	295	3	US-08-463-772-2	Sequence 2, Appl1
35	29	59.2	364	2	US-08-444-646-3	Sequence 3, Appl1
36	29	59.2	382	3	US-08-582-740-70	Sequence 70, Appl1
37	29	59.2	406	3	US-08-582-740-68	Sequence 68, Appl1
38	29	59.2	535	4	US-08-687-590-25	Sequence 25, Appl1
39	28	57.1	9	2	US-08-928-081-3	Sequence 3, Appl1
40	28	57.1	77	2	US-08-343-443B-12	Sequence 12, Appl1
41	28	57.1	77	2	US-08-343-443B-13	Sequence 13, Appl1
42	28	57.1	110	5	PCT-US96-03916-4	Sequence 4, Appl1
43	28	57.1	236	2	US-08-464-517-22	Sequence 22, Appl1
44	28	57.1	236	2	US-08-246-361A-22	Sequence 22, Appl1
45	28	57.1	236	3	US-08-463-772-22	Sequence 22, Appl1

ALIGNMENTS

RESULT 1
US-08-169-927-2
Sequence 2, Application US/08169927
Patent No. 5783441
GENERAL INFORMATION:
APPLICANT: Carl, Mitchell
APPLICANT: Dobson, Michael E.
APPLICANT: Ching, Wei Mei
APPLICANT: Dasch, Gregory A.
TITLE OF INVENTION: Gene and Protein Applicable to the
Preparation of Vaccines for Rickettsia typhi and
Rickettsia typhi and the Detection of Both
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical R & D Command
STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,927
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1612 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-927-2
Query Match 98.0% Score 48: DB 1: Length 1612:
Best Local Similarity 90.0% Pred. No. 0.14:

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TFIGAIAVDT 10
|:|||||
Db 1265 TFIGAIAVDT 1274

RESULT 2
US-08-738-172-4
; Sequence 4, Application US/08738172
; Patent No. 5939257
; GENERAL INFORMATION:
; APPLICANT: Szasz, Joseph
; APPLICANT: Davis, Maria
; TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,172
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/005,965
; FILING DATE: October 27, 1995
; APPLICATION NUMBER: 08/465,003
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/575,354
; FILING DATE: December 20, 1995
; APPLICATION NUMBER: 08/240,158
; FILING DATE: May 10, 1994
; APPLICATION NUMBER: 08/229,329
; FILING DATE: April 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 222/158
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEFAX: 67-3510
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 455 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-738-172-4

Query Match 69.4%; Score 34; DB 2; Length 455;
Best Local Similarity 70.0%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TFIGAIAVDT 10
|:|||||
Db 92 TYNCAIAVDT 101

RESULT 3

US-07-866-979-6
; Sequence 6, Application US/07866979
; Patent No. 5532347
; GENERAL INFORMATION:
; APPLICANT: Cone, Roger D
; APPLICANT: Mountjoy, Kathleen G
; TITLE OF INVENTION: Melanocyte Stimulating Hormone Receptor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Wilcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/866,979
; FILING DATE: 19920410
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5532347nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 92,154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEFAX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-866-979-6

Query Match 65.3%; Score 32; DB 1; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGAIAVD 9
|:|||||
Db 134 FIGAIAVD 141

RESULT 4
US-08-671-525B-2
; Sequence 2, Application US/08671525B
; Patent No. 5703220
; GENERAL INFORMATION:
; APPLICANT: Yamada, Tadataka
; APPLICANT: Gantz, Ira
; TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: US
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/671,525B
FILING DATE: June 27, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-000853DVB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810)641-1600
TELEFAX: (810)641-0270
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-671-525B-2

Query Match 65.3%; Score 32; DB 1; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGA1ATD 9
1:|||||
DB 134 FIGA1AVD 141

RESULT 5
US-08-672-109B-2
Sequence 2, Application US/08672109B
PATENT INFORMATION:
PATENT NO. 5710265
GENERAL INFORMATION:
APPLICANT: Yamada, Tadataka
APPLICANT: Gantz, Ira
TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: US
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/672,109B
FILING DATE: June 27, 1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-000853DVC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810)641-1600
TELEFAX: (810)641-0270
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-672-109B-2

Query Match 65.3%; Score 32; DB 1; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGA1ATD 9
1:|||||
DB 134 FIGA1AVD 141

RESULT 6
US-08-842-045-2
Sequence 2, Application US/08842045
PATENT INFORMATION:
PATENT NO. 581787
GENERAL INFORMATION:
APPLICANT: Yamada, Tadataka
APPLICANT: Gantz, Ira
TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: US
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/842,045
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-000853DVE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810)641-1600
TELEFAX: (810)641-0270
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-842-045-2

Query Match 65.3%; Score 32; DB 2; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGA1ATD 9
1:|||||
DB 134 FIGA1AVD 141

RESULT 7
US-08-466-906B-6
Sequence 6, Application US/08466906B
PATENT INFORMATION:
PATENT NO. 5849871
GENERAL INFORMATION:
APPLICANT: Cone, Roger D
APPLICANT: Mountjoy, Kathleen G
TITLE OF INVENTION: Melanocyte Stimulating Hormone Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,906B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: No. 5849871nan, Kevin E
REGISTRATION NUMBER: 35,303
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-906B-6

Query Match 65.3%; Score 32; DB 2; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGATATD 9
1:|||||1
Db 134 FLGATAVD 141

RESULT 8
US-08-842-238-2
Sequence 2, Application US/08842238
Patent No. 5869257
GENERAL INFORMATION:
APPLICANT: Yamada, Tadataka
APPLICANT: Gantze, Ira
TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: US
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/842,238
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-000853DVD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810)641-1600
TELEFAX: (810)641-0270
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-842-238-2

Query Match 65.3%; Score 32; DB 2; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGATATD 9
1:|||||1
Db 134 FLGATAVD 141

RESULT 9
US-08-780-749A-4
Sequence 4, Application US/08780749A
Patent No. 5932779
GENERAL INFORMATION:
APPLICANT: Lee, Frank
APPLICANT: Huszar, Dennis
APPLICANT: Gu, Wei
TITLE OF INVENTION: SCREENING METHODS FOR COMPOUNDS
USEFUL IN THE REGULATION OF BODY WEIGHT
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/780,749A
FILING DATE: 08-JAN-1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-064
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-780-749A-4

Query Match 65.3%; Score 32; DB 2; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGATATD 9
1:|||||1
Db 134 FLGATAVD 141

RESULT 10
US-08-706-281A-6
Sequence 6, Application US/08706281A
Patent No. 6100048
GENERAL INFORMATION:
APPLICANT: Cone, Roger D
APPLICANT: Fan, Wei
APPLICANT: Boston, Bruce A

APPLICANT: Kesterton, Robert A
APPLICANT: Lu, Dongxi
TITLE OF INVENTION: Methods and Reagents for Discovering and
TITLE OF INVENTION: Using Mammalian Melanocortin Receptor Agonists and Antagonists
TITLE OF INVENTION: To Modulate Feeding Behavior in Animals
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,281A
FILING DATE: 04-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 6100048nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 96,886
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-706-281A-6

Query Match 65.3%; Score 32; DB 3; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGA1ATD 9
DB 134 FIGA1AVD 141

RESULT 11
US-08-629-335B-2
Sequence 2, Application US/08629335B
Patent No. 6117975
GENERAL INFORMATION:
APPLICANT: Yamada, Tadataka
APPLICANT: Gantz, Ira
TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: US
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/629,335B
FILING DATE: July 23, 1996

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, DeAnn F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-000853DVA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810)641-1600
TELEFAX: (810)641-0270
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-629-335B-2

Query Match 65.3%; Score 32; DB 3; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGA1ATD 9
DB 134 FIGA1AVD 141

RESULT 12
US-09-201-746-6
Sequence 6, Application US/09201746
Patent No. 6268221
GENERAL INFORMATION:
APPLICANT: Cone, Roger D
APPLICANT: Mounjoy, Kathleen G
TITLE OF INVENTION: Melanocyte Stimulating Hormone Receptor
TITLE OF INVENTION: and uses
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201,746
FILING DATE: 01-DEC-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 6268221nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 92,154-J
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-201-746-6

Query Match 65.3%; Score 32; DB 4; Length 317;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGA1ATD 9
1:|||||
Db 134 FIGA1AVD 141

RESULT 13
US-08-911-319A-1
; Sequence 1, Application US/08911319A
; Patent No. 5968798
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN GLUTAREDOXIN BETA
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911.319A
; FILING DATE: August 14, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Muenzen, Colette C.
; REGISTRATION NUMBER: 39,784
; REFERENCE/DOCKET NUMBER: PF-0363 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-845-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 164 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: THP1NOT03
; CLONE: 2447829
; US-08-911-319A-1

*Query Match 62.2%; Score 30.5; DB 2; Length 164;
Best Local Similarity 80.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 FIGA1ATDT 10
1:|||||
Db 129 FIGG-ATDT 137

RESULT 14
US-09-352-619-1
; Sequence 1, Application US/09352619
; Patent No. 6084070
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN GLUTAREDOXIN BETA
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/352.619
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911.319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Muenzen, Colette C.
REGISTRATION NUMBER: 39,784
REFERENCE/DOCKET NUMBER: PF-0363 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-845-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 164 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: THP1NOT03
CLONE: 2447829
US-09-352-619-1

Query Match 62.2%; Score 30.5; DB 3; Length 164;
Best Local Similarity 80.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 FIGA1ATDT 10
1:|||||
Db 129 FIGG-ATDT 137

RESULT 15
US-08-973-461A-4
; Sequence 4, Application US/08973461A
; Patent No. 5989875
; GENERAL INFORMATION:
; APPLICANT: KOJIMA, HIROYUKI
; APPLICANT: OGAWA, YURI
; APPLICANT: KAWAMURA, KAZUE
; APPLICANT: SANO, KONOSUKE
; TITLE OF INVENTION: METHOD OF L-LYSINE BY FERMENTATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973.461A
; FILING DATE: 20-APR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-146054

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 0010-0901-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 292
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-973-461A-4

Query Match 61.2%; Score 30; DB 2; Length 292;
Best Local Similarity 62.5%; Pred. No. 95;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 IGAIAATDT 10
: 1 : ||||
Db 259 LGIVATDT 266

Search completed: September 28, 2001, 12:27:45
Job time: 5469 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2001, 10:26:15 : Search time 52.82 Seconds
(without alignments)
14.422 Million cell updates/sec

Title: US-09-551-645-1
Perfect score: 49
Sequence: 1 TFGATATDT 10

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	1645	2 JN0896	crystalline surfac
2	48	98.0	1643	2 D71630	outer membrane pro
3	40	81.6	520	2 D70035	permease homolog y
4	39	79.6	1300	2 S07575	outer membrane pro
5	39	79.6	1651	2 JC1340	outer membrane pro
6	37	75.5	344	2 H70030	conserved hypotet
7	36	73.5	174	2 A45937	early chorion prot
8	35	71.4	430	2 A84887	probable phosphat
9	35	71.4	2434	2 S44861	DNA topoisomerase
10	34	69.4	399	2 F70937	hypothetical prote
11	34	69.4	468	2 D72097	amino acid antipor
12	34	69.4	468	2 B86526	amino acid transp
13	33	67.3	179	2 T00893	thioredoxin F21B7.
14	33	67.3	179	2 A64846	probable membrane
15	33	67.3	179	2 B85665	hypothetical prote
16	33	67.3	292	2 S65362	o-pyrocatechuate d
17	33	67.3	299	2 T18563	hypothetical prote
18	33	67.3	314	2 T25842	hypothetical prote
19	33	67.3	385	2 C83174	hypothetical prote
20	33	67.3	389	2 T43979	hypothetical prote
21	33	67.3	531	2 T14640	cytochrome P450 cy
22	33	67.3	555	2 T45351	ferredoxin import
23	33	67.3	2529	2 B64635	toxin-like outer m
24	32	65.3	140	2 S61385	lcmx protein - Leg
25	32	65.3	169	2 B23548	chorion class CB p
26	32	65.3	199	2 A19324	conserved hypotet
27	32	65.3	227	2 E69079	ribonuclease III U
28	32	65.3	236	2 F82919	hypothetical prote
29	32	65.3	271	2 E83546	

30	32	65.3	388	2 F84295	hypothetical prote
31	32	65.3	317	2 T12055	melanocyte stimula
32	32	65.3	317	2 S29204	melanotropin recep
33	32	65.3	317	2 S45708	MSH receptor - bov
34	32	65.3	325	2 T22994	hypothetical prote
35	32	65.3	442	2 E82980	probable oxidoredu
36	32	65.3	448	2 D75197	tRNA nucleotidyltr
37	32	65.3	454	2 D86793	drug-export protei
38	32	65.3	466	2 H81697	amino acid antipor
39	32	65.3	481	2 C71230	hypothetical prote
40	32	65.3	540	2 F65135	phosphoenolpyruvat
41	32	65.3	542	2 D86005	phosphoenolpyruvat
42	32	65.3	540	2 B82039	phosphoenolpyruvat
43	32	65.3	799	2 B86735	hypothetical prote
44	32	65.3	819	2 T48307	hypothetical prote
45	32	65.3	929	2 C84493	probable Athlia re

ALIGNMENTS

RESULT 1
JN0896
C:Species: Rickettsia typhi
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #ext_change 28-May-1999
C:Accession: JN0896; PNO686
R:Rhm, M.J.; Kim, K.K.; Kim, I.; Chang, W.H.
Gene 133, 129-133, 1993
A:Title: Cloning and sequence analysis of the gene encoding the crystalline surface 1
A:Reference number: JN0896; MID:94040787
A:Accession: JN0896
A:Molecule type: DNA
A:Residues: 1-1645 <HA>
A:Cross-references: GB:104661; NID:9152474; PIDN:AA848987.1; PID:91871223
A:Molecule type: protein
A:Residues: 1353-1371 <HA2>
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-1645/Product: crystalline surface layer protein #status predicted <MAT>

Query Match 100.0%; Score 49; DB 2; Length 1645;
Best Local Similarity 100.0%; Pred. No. 0.09;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TFGATATDT 10
DB 1296 TFGATATDT 1305

RESULT 2
D71630
N:Alternate names: Cell surface antigen scas; paracrystalline surface-layer protein;
N:Contents: 32k beta peptide
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #ext_change 03-Nov-2000
C:Accession: D71630; A36473
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MID:99039499
A:Accession: D71630
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1643 <AND>
A:Cross-references: GB:AJ235273; GB:AJ235269; NID:93861237; PIDN:CA15140.1; PID:el34
A:Experimental source: strain Madrid E
R:Carl, M.; Dobson, M.E.; Chang, W.M.; Dasch, G.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 8237-8241, 1990
A:Title: Characterization of the gene encoding the protective paracrystalline-surface
A:Reference number: A36473; MID:91045972

A:Accession: A36473
 A:Molecule type: DNA
 A:Residues: 33-177, 'VC', 180-190, 'INRSRSSYHLVS', 202-211, 'I', 213-256, 'A', 258-312, 'L', 314-315
 A:Cross-references: GB:M3647; NID:9152497; PIDN:AA26390.1; PID:9152498
 A:Note: the authors translated the codon AAT for residue 146 as Ala, CGT for residue 478
 5 is inconsistent with the translation of the nucleotide sequence
 A:Note: parts of this sequence were determined by protein sequencing
 R:Hackstadt, T.; Messer, R.; Cleplak, W.; Peacock, M.G.
 Infect. Immun. 60, 159-165, 1992
 A:Title: Evidence for proteolytic cleavage of the 120-kilodalton outer membrane protein
 A:Reference number: A43869; MUID:92104668
 A:Contents: unannotated
 A:Note: experimental evidence of proteolytic processing
 C:Genetics:
 A:Gene: ompB; sca5; RP704; spap
 C:Keywords: glycoprotein; surface antigen
 F:2-135/Product: outer membrane protein B #status predicted <MAT>
 F:1353-1643/Product: 32k beta peptide #status experimental <MAT>

Query Match 98.0%; Score 48; DB 2; Length 1643;
 Best Local Similarity 90.0%; Pred. No. 0.14;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TFIGAATDT 10
 ||:|||||
 Db 1296 TFIGAATDT 1305

RESULT 3
 D70035
 permease homolog yveA - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: D70035

R:Kunst, F.; Ogatawa, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall
 lech, J.; Harwood, C.R.; Henuit, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koester, P.; Konlingstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadle, Y.; Sato, T.; Scanlon
 A:Authors: Schell, S.; Schotter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
 akuch, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: D70035
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-520 <KUN>
 A:Cross-references: GB:299121; GB:AL009126; NID:92653827; PIDN:CAB15452.1; PID:92653960
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yveA

Query Match 81.6%; Score 40; DB 2; Length 520;
 Best Local Similarity 80.0%; Pred. No. 2;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TFIGAATDT 10
 |||||:|
 Db 253 TFIGAATDT 262

RESULT 4
 S07575
 outer membrane protein B - Rickettsia rickettsii

N:Alternate names: 120K surface-exposed protein
 N:Contains: 32k beta peptide
 C:Species: Rickettsia rickettsii
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Dec-1998
 C:Accession: S07575; S18579; A43869
 R:Gillmore, R.D.; Joste, N.; McDonald, G.A.
 Mol. Microbiol. 3, 1579-1586, 1989

A:Title: Cloning, expression and sequence analysis of the gene encoding the 120kd sur
 A:Reference number: S07575; MUID:90136087
 A:Accession: S07575

A:Molecule type: DNA
 A:Residues: 1-994, 'DLKLEH', 1001, 'GS', 1005-1300 <GILL>

A:Cross-references: EMBL:X16353
 A:Note: this sequence has been revised in reference S18579

R:Gillmore, R.D.; Joste, N.; McDonald, G.A.
 Mol. Microbiol. 5, 3089, 1991

A:Reference number: S18579; MUID:92236427
 A:Contents: erratum

A:Accession: S18579

A:Molecule type: mRNA

A:Residues: 994-1004 <GILL2>
 R:Hackstadt, T.; Messer, R.; Cleplak, W.; Peacock, M.G.

Infect. Immun. 60, 159-165, 1992

A:Title: Evidence for proteolytic cleavage of the 120-kilodalton outer membrane prote
 A:Reference number: A43869; MUID:92104668

A:Accession: A43869

A:Molecule type: protein
 A:Residues: 1008-1027 <HAC>

A:Note: sequence extracted from NCBI backbone (NCBIP:74388) and corrected to correspo
 C:Keywords: glycoprotein
 F:2-1007/Product: outer membrane protein B #status predicted <MAT>
 F:1008-1300/Product: 32k beta peptide #status experimental <MAT>

Query Match 79.6%; Score 39; DB 2; Length 1300;
 Best Local Similarity 77.8%; Pred. No. 8.5;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 TFIGAATDT 10
 ||:|||||
 Db 948 TFIGAATDT 956

RESULT 5

JC1340
 outer membrane protein B precursor - Rickettsia japonica

C:Species: Rickettsia japonica
 C:Date: 30-Sep-1993 #sequence_revision 21-Jan-1997 #text_change 13-Nov-1998
 C:Accession: JC1340

R:Ryan, Y.; Uchida, T.
 Chinese J. Microbiol. Immunol. 16, 220-226, 1996

A:Title: Cloning and sequence analysis of the gene encoding the 120kDa outer membrane
 A:Reference number: JC1340
 A:Accession: JC1340

A:Molecule type: DNA
 A:Residues: 1-1651 <VAN>

A:Note: The authors translated the codon GGT for residue 241 as Cys, CAA for residue
 C:Genetics:
 A:Gene: ompB

C:Keywords: membrane protein
 F:1-134/Domains: signal sequence #status predicted <SIG>
 F:35-1651/Product: outer membrane protein B #status predicted <MAT>

Query Match 79.6%; Score 39; DB 2; Length 1651;
 Best Local Similarity 77.8%; Pred. No. 11;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 TFIGAATDT 10
 ||:|||||
 Db 1302 TFIGAATDT 1310

```

RESULT 6
H70030
conserved hypothetical protein yvbx - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: H70030
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Beren
C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.: Erlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
leach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinios,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
R:Authors: Schleich, S.; Schoeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: H70030
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-344 <KUN>
A:Cross-references: GB:299121; GB:AL009126; NID:g2635827; PID:CA015407.1; PID:g2635915
A:Experimental source: strain 168
C:Genetics:
A:Gene: yvbx
C:Superfamily: hypothetical protein yvbx

Query Match 75.5%; Score 37; DB 2; Length 344;
Best Local Similarity 70.0%; Pred. No. 5.6;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 TFIGATATDT 10
I:|||||
Db 50 TYWMAIATDT 59

RESULT 7
A45937
early chorion protein 5H4 precursor - silkworm
C:Species: Bombyx mori (silkworm)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 20-Aug-1999
C:Accession: A45937
R:Hilber, B.L.; Burke, W.D.; Lecandou, R.; Rodakis, G.C.; Eickbush, T.H.
Dev. Biol. 125, 423-431, 1988
A:Title: Organization and expression of three genes from the silkworm early chorion locu
A:Reference number: A45937; MUID:88112521
A:Accession: A45937
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-174 <HIB>
A:Cross-references: GB:M19075; NID:g155973; PIDN:AAA27830.1; PID:g155977
C:Superfamily: chorion class A protein pc292

Query Match 73.5%; Score 36; DB 2; Length 174;
Best Local Similarity 75.0%; Pred. No. 4.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 FIGATATD 9
I:|||||
Db 110 FIGAVATD 117

RESULT 8
A4887
probable phosphatidate cytidyltransferase [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

```

```

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A4887
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A48420; MUID:20083487
A:Accession: A4887
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-430 <STO>
A:Cross-references: GB:AE002093; NID:g4895237; PIDN:AAD32822.1; GSPDB:GN00139
C:Genetics:
A:Gene: AT9G45150
A:Map position: 2

Query Match 71.4%; Score 35; DB 2; Length 430;
Best Local Similarity 70.0%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 TFIGATATDT 10
I:|||||
Db 301 SFGVIATDT 310

RESULT 9
S44861
DNA topoisomerase II - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 23-Mar-2001
C:Accession: S44861
R:Wilson, R.
submitted to the EMBL Data Library, December 1992
A:Description: Sequence of the C. elegans cosmid R05D3.
A:Reference number: S44733
A:Accession: S44861
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2434 <WIL>
A:Cross-references: EMBL:L07144; NID:g156412; PID:g156413
C:Genetics:
A:Introns: 32/3; 128/3; 361/3; 506/1; 710/3; 1000/2; 1045/3; 1221/2; 1294/1; 1378/2;

Query Match 71.4%; Score 35; DB 2; Length 2434;
Best Local Similarity 66.7%; Pred. No. 11e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 TFIGATATD 9
I:|||||
Db 1070 TYVGATITD 1078

RESULT 10
F70937
hypothetical protein RV2188c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70937
R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
J.; Connor, M.A.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: F70937
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-399 <COL>
A:Cross-references: GB:AL021957; GB:AL123456; NID:g3242293; PIDN:CA17492.1; PID:g291

```

A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV2188c
 C:Superfamily: probable hexosyltransferase ytxN

Query Match 69.4%; Score 34; DB 2; Length 399;
 Best Local Similarity 66.7%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TFIGAATD 9
 ||| : |||
 Db 271 TFIGAATD 279

RESULT 11
 D72097
 amino acid antiporter CP0476 [Imported] - Chlamydomophila pneumoniae (strains CWL029 and A
 N:Alternate names: amino acid transporter
 C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: D72097; A81573
 R:Kakuman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: D72097
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-468 <ARN>
 A:Cross-references: GB:AE001613; GB:AE001363; NID:94376550; PIDN:AAD18431.1; PID:9437655
 A:Experimental source: strain CWL029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Koloney, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255
 A:Accession: A81573
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-468 <REA>
 A:Cross-references: GB:AE002209; GB:AE002161; NID:97189393; PIDN:AAF38309.1; PID:9718939
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: xasa; CP0476
 C:Superfamily: L-lysine transport protein

Query Match 69.4%; Score 34; DB 2; Length 468;
 Best Local Similarity 88.9%; Pred. No. 33;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 FIGAATD 10
 ||||| : |||
 Db 233 FIGAATD 241

RESULT 12
 B86526
 amino acid transporter [Imported] - Chlamydomophila pneumoniae (strain J138)
 C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
 C:Accession: B86526
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tsubuchi, M.; Kishl, F.; Ouchi, K.; Shiba, T.; Is
 Nucleic Acids Res. 28, 2311-2314, 2000
 A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A:Reference number: A6491; MUID:20330349
 A:Accession: B86526
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-468 <STO>
 A:Cross-references: GB:BA000008; NID:98978656; PIDN:BAA98492.1; GSPDB:GN00142
 A:Experimental source: strain J138

C:Genetics:
 A:Gene: xasa

Query Match 69.4%; Score 34; DB 2; Length 468;
 Best Local Similarity 88.9%; Pred. No. 33;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 FIGAATD 10
 ||||| : |||
 Db 233 FIGAATD 241

RESULT 13
 T00893
 thioresoxin F21B7.7 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 05-May-2000
 C:Accession: T00893
 R:Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.;
 eologists, A.; Ecker, J.R.
 Submitted to the EMBL Data Library, January 1998
 A:Description: Genomic sequence for Arabidopsis thaliana BAC F21B7.
 A:Reference number: Z14208
 A:Accession: T00893
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-179 <SHI>
 A:Cross-references: EMBL:AC002560; NID:92618677; PIDN:AAB97698.1; PID:92809238; GSPDB
 C:Genetics:
 A:Gene: ATSP:F21B7.7
 A:Map position: 1
 A:Introns: 76/1
 C:Superfamily: thioresoxin; thioresoxin homology
 C:Keywords: redox-active disulfide
 F:82-165/Domain: thioresoxin homology <TXN>
 F:104-107/Disulfide bonds: redox-active #status predicted

Query Match 67.3%; Score 33; DB 2; Length 179;
 Best Local Similarity 60.0%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TFIGAATD 10
 ||||| : |||
 Db 161 TFIGAATD 170

RESULT 14
 A64846
 probable membrane protein ycdZ - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
 C:Accession: A64846
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: A64846
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-179 <BLAT>
 A:Cross-references: GB:AE000205; GB:U00096; NID:91787265; PIDN:AAC74120.1; PID:917872
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: ycdZ
 C:Keywords: transmembrane protein
 F:19-35/Domain: transmembrane #status predicted <TM1>
 F:37-53/Domain: transmembrane #status predicted <TM2>
 F:68-84/Domain: transmembrane #status predicted <TM3>
 F:94-110/Domain: transmembrane #status predicted <TM4>
 F:115-131/Domain: transmembrane #status predicted <TM5>

F:138-154/Domain: transmembrane #status predicted <TM6>

Query Match 67.3%; Score 33; DB 2; Length 179;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TFIGATAT 8
|||||
Db 123 TFIGACAT 130

RESULT 15

B85665

hypothetical protein ycdz [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 31-Mar-2001

C:Accession: B85665

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: B85665

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-179 <STO>

A:Cross-References: GB:AE005174; NID:g12514565; PIDN:AAG55782.1; GSPDB:GN00145; UWGP:216

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ycdz

Query Match 67.3%; Score 33; DB 2; Length 179;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TFIGATAT 8
|||||
Db 123 TFIGACAT 130

Search completed: September 28, 2001, 12:27:10
Job time: 7255 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2001, 12:26:05 ; Search time 53.45 Seconds
(without alignments)
6.409 Million cell updates/sec

Title: US-09-551-645-1
Perfect score: 49
Sequence: 1 TFIGATATDR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	49	100.0	1 OMPB_RICTY	P96989 r outer mem
2	48	98.0	2 OMPB_RICPY	O53020 r outer mem
3	39	79.6	1300 1	P14914 rickettsia
4	39	79.6	1654 1	O53047 r outer mem
5	36	72.5	174 1	P08830 bomblyx mori
6	35	71.4	1254 1	P34544 caenorhabditis
7	35	71.4	1656 1	O06653 r outer mem
8	33	67.3	106 1	O48920 salmonella
9	33	67.3	179 1	P75916 escherichia
10	33	67.3	179 1	P80346 aspergillus
11	33	67.3	292 1	O99499 emericella
12	33	67.3	531 1	P08830 antheraea p
13	33	67.3	553 1	P80346 aspergillus
14	32	65.3	167 1	O48920 salmonella
15	32	65.3	317 1	P75916 escherichia
16	32	65.3	317 1	P80346 aspergillus
17	32	65.3	317 1	O99499 emericella
18	32	65.3	317 1	P08830 antheraea p
19	32	65.3	317 1	P80346 aspergillus
20	32	65.3	317 1	O99499 emericella
21	32	65.3	317 1	P08830 antheraea p
22	32	65.3	317 1	P80346 aspergillus
23	32	65.3	317 1	O99499 emericella
24	32	65.3	317 1	P08830 antheraea p
25	32	65.3	317 1	P80346 aspergillus
26	32	65.3	317 1	O99499 emericella
27	32	65.3	317 1	P08830 antheraea p
28	31	63.3	317 1	P80346 aspergillus
29	31	63.3	362 1	P56444 capra hircu
30	31	63.3	392 1	P56444 capra hircu
31	31	63.3	392 1	P56444 capra hircu
32	31	63.3	427 1	P56444 capra hircu
33	31	63.3	716 1	P56444 capra hircu

ALIGNMENTS

RESULT 1
OMP_RICTY STANDARD; PRT; 1645 AA.

AC OMPB_RICTY
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (POMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).
OS SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxId=785;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT Cloning and sequence analysis of the gene encoding the crystalline
RT surface layer protein of Rickettsia typhi.
RL Gene 133:129-133(1993)
RN [2]
RP STRAIN: WILMINGTON.
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT Mapping of monoclonal antibody binding sites on CNBR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.
RL Mol. Immunol. 29:95-105(1992).
RN [3]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplik W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of Rickettsia: identification of an avirulent
RT mutant deficient in processing."
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
CC AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch)

```

DR EMBL: L04661; AAB48987.1; -.
KW Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.
FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
FT CONFLICT 657 657 H -> N (IN REF. 2).
FT CONFLICT 842 842 V -> I (IN REF. 2).
FT CONFLICT 1071 1071 G -> A (IN REF. 2).
FT CONFLICT 1306 1306 G -> S (IN REF. 2).
SO SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 100.0%; Score 49; DB 1; Length 1645;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TFICAIATDT 10
Db 1296 TFEICAIATDT 1305

RESULT 2
OMPBR_RICPR STANDARD: PRT; 1643 AA.
AC 053020; Q9ZCM0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB OR SPAP OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID:782;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BREINL;
RA MEDLINE=91045972; PubMed=2122457;
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
RT surface-layer protein of Rickettsia prowazekii: presence of a
RT truncated identical homolog in Rickettsia typhi."
RT Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BREINL;
RA Moron C.G., Yu X.J., Walker D.H.;
RT "Sequence analysis of ompB of Rickettsia prowazekii."
RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RA MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Slicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RT Nature 396:133-140(1998).
RN [4]
RP PARTIAL SEQUENCE.
RC STRAIN-BREINL;
RA MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii."
RT Mol. Immunol. 29:95-105(1992).
RN [5]
RP IDENTIFICATION OF CLEAVAGE SITE.

```

```

RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplak W. Jr., Paacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing."
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37647; AAA26390.1; ALT_INIT.
DR EMBL: AF161079; AAD42234.1; -.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
FT VARIANT 257 257 Y -> A (IN STRAIN BREINL).
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
FT CONFLICT 178 179 AA -> VC (IN REF. 1)
FT CONFLICT 191 201 TTQAPLPLGA -> INRSSTYLVS (IN REF. 1).
FT CONFLICT 212 212 T -> I (IN REF. 1).
FT CONFLICT 313 313 O -> L (IN REF. 1).
FT CONFLICT 1104 1104 D -> G (IN REF. 2).
FT CONFLICT 1123 1123 T -> S (IN REF. 2).
SO SEQUENCE 1643 AA; 169854 MW; 735FDF392B6346CC CRC64;

Query Match 98.0%; Score 48; DB 1; Length 1643;
Best Local Similarity 90.0%; Pred. No. 0.077;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TFEICAIATDT 10
Db 1296 TFEICAIATDT 1305

RESULT 3
120K_RICRI STANDARD: PRT; 1300 AA.
ID 120K_RICRI
AC P14914;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 120 KDA SURFACE-EXPOSED PROTEIN.
GN P120.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RA MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii."
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS
CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.

```

```
CC -I- DISEASE: RICKETTSIA RICKETTSII IS THE CAUSATIVE AGENT OF ROCKY
CC MOUNTAIN SPOTTED FEVER (RMSF).
CC -I- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPTIOPES
CC CONFERRING ANTIGENICITY TO THE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16353; CAA34402.1; -.
DR PIR: S07575; S07575.
KW Antigen; Glycoprotein; S-layer.
FT CARBOHYD 7 N-LINKED (POTENTIAL).
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SO SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;

Query Match          79.6%; Score 39; DB 1; Length 1300;
Best Local Similarity 77.8%; Pred. No. 4.2;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY      2 FIGAINDT 10
        I::III III
.Db     948 FVGAIIVDT 956

RESULT   4
OMP_B_RICRI STANDARD; PTR: 1654 AA.
AC Q53047;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB.
OS Rickettsia rickettsii.
```

CC Bacteriophage: Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsia; Rickettsia.
 CC NCBI_Taxid=783;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=R;
 RC MEDLINE=90136087; PubMed=2515418;
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
 RT "Cloning, expression and sequence analysis of the gene encoding the
 RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
 RL Mol. Microbiol. 3:1579-1586(1989).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X16353; CAA34403.1; -.
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1334 1634 32 KDA BETA PEPTIDE.
 FT DOMAIN 1181 1188 POLY-THR.
 SQ SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087P618 CRC64;
 OY 2 FIGATATDT 10 79.6%; Score 39; DB 1; Length 1654;
 1:111111 Pred. No. 5.4;
 DB 1302 FVGATVDT 1310 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 RESULT 5
 CHGB_BOMMO STANDARD; PRT; 174 AA.
 AC P08830;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-FEB-1991 (Rel. 17, Last annotation update)
 DE CHORION CLASS CB PROTEIN M5H4 PRECURSOR.
 OS Bombyx mori (Silk moth).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 CC Bombycoidea; Bombycidae; Bombyx.
 CC NCBI_Taxid=7091;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=08112521; PubMed=3338621;
 RA Hübner B.L., Burke W.D., Lecanidou R., Rodakis G.C., Eickbush T.H.;
 RT "Organization and expression of three genes from the silkworm early
 RT chorion locus.";
 RL Dev. Biol. 125:423-431(1988).
 RL [2]
 RP SEQUENCE OF 6-174 FROM N.A.

```

RX MEDLINE=86313609; Pubmed=3462711;
RA Lecanidou R., Rodakis G.C., Eickbush T.H., Kafatos F.C.;
RT "Evolution of the silk moth chorion gene superfamily: gene families
CC CA and CB.";
CC Proc. Natl. Acad. Sci. U.S.A. 83:6514-6518(1986).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC SILK MOTH.
CC -1- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH
CC BELONG CLASSES B, CB AND HCB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M19075; AAA27830.1; -
DR EMBL; M13835; AAA27829.1; -
DR PIR; B23548; B23548.
DR PIR; A45937; A45937.
DR HSP; P00586; 1DTP.
DR InterPro; IPR002635; -
DR Pfam; PF01723; Chorion; 1.
KW Eggshell; Chorion; Repeat; Multigene family; Signal.
FT SIGNAL 1 20
FT CHAIN 21 174 CHORION CLASS CB PROTEIN M5H4.
FT DOMAIN 21 71 LEFT ARM.
FT DOMAIN 72 142 CENTRAL DOMAIN.
FT DOMAIN 143 174 RIGHT ARM.
FT CONFLICT 115 115 A -> V (IN REF. 2).
SQ SEQUENCE 174 AA; 17000 MW; B0ED1BC8928CC568 CRC64;

Query Match 73.5%; Score 36; DB 1; Length 174;
Best Local Similarity 75.0%; Pred. No. 2.3;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 FIGA1ATD 9
   1:|||||
DB 110 FLGAVATD 117

RESULT 6
YNCA CAEEL STANDARD; PRT: 1254 AA.
AC P34544;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHEICAL 142.5 KDA PROTEIN R05D3.11 IN CHROMOSOME III.
GN R05D3.11.
OS Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
CC Rhabditidae; Pelodermidae; Caenorhabditis.
CC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RX MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Anscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sultson J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Wohldman P.;
RP "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
```

```

RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L07144; - NOT_ANNOTATED_CDS.
DR WormPep; R05D3.11; CE00529.
DR InterPro; IPR001214; -
DR InterPro; IPR001739; -
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00856; SET; 1.
DR ProSite; PS0280; SET; 1.
KW Hypothetical protein.
SQ SEQUENCE 1254 AA; 142489 MW; E06789748CB6D9FC CRC64;

Query Match 71.4%; Score 35; DB 1; Length 1254;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 TFIGA1ATD 9
   1:|||||
DB 1103 TYVGALILD 1111

RESULT 7
OMP8_RICJA STANDARD; PRT: 1656 AA.
ID OMP8_RICJA
AC 006653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE [CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
GN OMPB.
OS Rickettsia japonica.
CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Rickettsiae; Rickettsia.
CC NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB003681; BAA20138.1; -
```

KM Antigen: S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1 1338 32 KDA BETA PEPTIDE.
 FT CHAIN 1339 1656 POLY-GLY.
 FT DOMAIN 528 533
 SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match
 Best Local Similarity 71.4%; Score 35; DB 1; Length 1656;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 FIGAIAIT 10
 DB 1307 FVGTIVDT 1315

RESULT 8
 YCDZ_SALTY STANDARD: PRT; 106 AA.
 AC 054290;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN IN PHOH-CSGG INTERGENIC REGION (FRAGMENT).
 GN YCDZ.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SR-11;
 RX MEDLINE-96117058; PubMed-9457880;
 RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
 RA "Curl" fibers are highly conserved between *Salmonella typhimurium* and
 RA *Escherichia coli* with respect to operon structure and regulation.";
 RL J. Bacteriol. 180:722-731(1998).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO E.COLI YABC.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ002301; CA05311.1; -
 DR StyGene; SG10739; ycdz.
 KM Hypothetical protein; Transmembrane.
 FT NON_TER 1 1
 FT TRANSMEM 3 23 POTENTIAL.
 FT TRANSMEM 24 44 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 69 89 POTENTIAL.
 SQ SEQUENCE 106 AA; 11318 MW; DFD057BC3565A2C5 CRC64;

Query Match
 Best Local Similarity 67.3%; Score 33; DB 1; Length 106;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TFIGAIAIT 8
 DB 53 TFIGACAT 60

RESULT 9
 THML_ARATH STANDARD: PRT; 179 AA.
 AC 048737;
 DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE THIOREDOXIN M-TYPE 1, CHLOROPLAST PRECURSOR (TRX-M1).
 GN F2187.28.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20047910; PubMed-10580150;
 RA Mestres-Ortega D., Meyer Y.;
 RT "The Arabidopsis thaliana genome encodes at least four thioredoxins m
 RT and a new prokaryotic-like thioredoxin.";
 RL Gene 240:307-316(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
 RA Shin P., Altafi H., Bel O., Chin C., Chlou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PARTICIPATES IN VARIOUS REDOX REACTIONS THROUGH THE
 CC REVERSIBLE OXIDATION OF THE ACTIVE CENTER DITHIOL, TO A DISULFIDE.
 CC THE M FORM IS KNOWN TO ACTIVATE NADP-MALATE DEHYDROGENASE (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY. PLANT M-TYPE.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF095749; AAF15948.1; -
 DR EMBL: AC002560; AAF6525.1; -
 DR HSSP: P00274; I77P.
 DR Mendel; 27907; Arabid.1238; 27907.
 DR InterPro; IPR000063; -
 DR Pfam; PF000085; thioRed; 1.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR PROSITE: PS00194; THIOREDOXIN; 1.
 KM Redox-active center; Electron transport; Chloroplast; Transit peptide;
 KW Multigene family.
 FT TRANSIT 1 1
 FT CHAIN 1 179 CHLOROPLAST (POTENTIAL).
 FT DISULFID 104 107 THIOREDOXIN M-TYPE 1.
 SQ SEQUENCE 179 AA; 19664 MW; 35B9E7C1D132F492 CRC64;

Query Match
 Best Local Similarity 67.3%; Score 33; DB 1; Length 179;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TFIGAIAIT 10
 DB 161 TFIGAVSKDT 170

RESULT 10
 YCDZ_ECOLI STANDARD: PRT; 179 AA.
 AC P75916;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

```

DE HYPOTHETICAL 18.8 KDA PROTEIN IN PHO-HCSGG INTERGENIC REGION.
GN YCD2.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426517; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Stampel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO E.COLI YACB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000205; AAC74120.1; -
DR EMBL: D90740; BAA35817.1; -
DR EMBL: D90741; BAA35826.1; -
DR EcGene; EG13872; ycd2.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 41 61 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 139 159 POTENTIAL.
SO SEQUENCE 179 AA; 18779 MW; 84C8B0A1B2BB1B1D CRC64;

Query Match 67.3%; Score 33; DB 1; Length 179;
Best Local Similarity 87.5%; Pred. NO. 9.9;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 FIGA1AT 8
DB 123 FIGACAT 130

```

```

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE.
RX MEDLINE=95324511; PubMed=7601088;
RA Santha R., Savithri H.S., Rao N.A., Valadanathan C.S.;
RT "2,3-Dihydroxybenzoic acid decarboxylase from Aspergillus niger. A
RT novel decarboxylase."
RL Eur. J. Biochem. 230:104-110(1995).
CC -1- CATALYTIC ACTIVITY: 2,3-DIHYDROXYBENZOATE + CO(2) ->
CC -1- PATHWAY: LAST STEP IN THE METABOLISM OF INDOLE TO CATÉCHOL.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- MISCELLANEOUS: THE MW OF THE COMPLETE PROTEIN IS ABOUT 38 KDA (345
CC RESIDUES).
KW Lysase; Decarboxylase.
FT ACT SITE 263 263
FT UNSURE 247 247 OR G.
FT UNSURE 250 250 OR G.
FT NON_CONS 32 33
FT NON_CONS 66 67
FT NON_CONS 88 89
FT NON_CONS 121 122
FT NON_CONS 150 151
FT NON_CONS 160 161
FT NON_CONS 170 171
FT NON_CONS 183 184
FT NON_CONS 194 195
FT NON_CONS 207 208
FT NON_CONS 214 215
FT NON_CONS 224 225
FT NON_CONS 229 230
FT NON_CONS 236 237
FT NON_CONS 240 241
FT NON_CONS 247 248
FT NON_CONS 255 256
FT NON_CONS 264 265
FT NON_TER 292 292
SO SEQUENCE 292 AA; 33000 MW; EF78074854BAD925 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 292;
Best Local Similarity 66.7%; Pred. NO. 16;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 2 FIGA1ATDT 10
DB 98 FIGALVNDT 106

RESULT 12
C7EL-SORBI
ID C7EL-SORBI STANDARD; PRT; 531 AA.
AC 048958;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME P450 71E1 (EC 1.14.-.-).
GN CYP71E1.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC Andropogoneae; Sorghum.
OX NCBI_TaxID=4558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV SS1000; TISSUE=Etolated seedling;
RX MEDLINE=96145474; PubMed=9484480;
RA Bak S., Kahn R.A., Nielsen H.L., Moeller B.L., Halkier B.A.;
RT "Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99
RT from Sorghum bicolor (L.) Moench by a PCR approach and identification
RT by expression in Escherichia coli of CYP71E1 as a multifunctional
RT cytochrome P450 in the biosynthesis of the cyanogenic glucoside

```

RT dhurrin.":
 CC Plant Mol. Biol. 36:393-405(1998).
 CC -1- FUNCTION: CATALYZES THE CONVERSION OF P-HYDROXYPHENYLACETALDOXIME
 CC TO P-HYDROXYMANDELONITRILE. THE DEHYDRATION OF THE OXIME TO THE
 CC CORRESPONDING NITRILE IS FOLLOWED BY A C-HYDROXYLATION OF THE
 CC NITRILE TO PRODUCE P-HYDROXYMANDELONITRILE.
 CC -1- PATHWAY: BIOSYNTHESIS OF THE CYANOGENIC GLUCOSIDE DHURBIN.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC
 CC EMBL: AF029858: AAC39318.1; -
 CC InterPro: IPR001128; -
 CC InterPro: IPR002401; -
 CC Pfam: PF00067; P450; 1.
 CC PRINTS: PR00463; EP4501.
 CC PROSITE: PS00086; CYTOCHROME P450; 1.
 CC Oxidoreductase; Monooxygenase; Transmembrane; Heme;
 CC Endoplasmic reticulum.
 CC TRANSMEM 18 38 POTENTIAL.
 CC FT BINDING 468 468 HEME (BY SIMILARITY).
 CC SEQUENCE 531 AA; 59088 MW; 2272E4AA910890D2 CRC64;
 CC
 CC Query Match 67.3%; Score 33; DB 1; Length 531;
 CC Best Local Similarity 70.0%; Pred. No. 29;
 CC Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC 1 TFGAINTD 10
 CC 325 TFGAIDTSS 334
 CC
 CC RESULT 13
 CC ID HIS5_EMENT STANDARD: PRT: 553 AA.
 CC AC 09P49;
 CC DT 01-OCT-2000 (Rel. 40; Created)
 CC DT 01-OCT-2000 (Rel. 40; Last sequence update)
 CC DT 01-OCT-2000 (Rel. 40; Last annotation update)
 CC DE BIFUNCTIONAL HISTIDINE BIOSYNTHESIS PROTEIN HIS5 [INCLUDES: HIS-
 CC AMIDOTRANSFERASE (EC 2.4.2.-); HISF-TYPE CYCLASE].
 CC GN HISF.
 CC OS Emericella nidulans (Aspergillus nidulans).
 CC CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 CC OX NCBI_TaxID=5072;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=4234;
 CC RA Valerius O., Draht O., Hoffmann B., Kuebler E., Braus G.H.;
 CC RT Isolation and characterization of the hisF gene of Aspergillus
 CC nidulans.
 CC RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES AN AMIDOTRANSFERASE REACTION THAT GENERATES
 CC IMIDAZOLE-GLYCEROL PHOSPHATE AND 5-AMINOIMIDAZOL-4-CARBOXYAMIDE
 CC RIBONUCLEOTIDE, WHICH IS USED FOR PURINE SYNTHESIS.
 CC -1- FUNCTION: CATALYZES THE CYCLIZATION REACTION THAT PRODUCES
 CC D-PRYTHRO-IMIDAZOLE GLYCEROL PHOSPHATE.
 CC -1- PATHWAY: FIFTH AND SIXTH STEPS IN HISTIDINE BIOSYNTHETIC PATHWAY.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE HISF FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HISA / HISF
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC
 CC EMBL: AF159463; AAF80376.1; -
 CC DR PROSITE: PS00442; GATASE_TYPE_1;
 CC KW Histidine biosynthesis; Transferase; Glutamine amidotransferase;
 CC Multifunctional enzyme.
 CC FT DOMAIN 1 211
 CC FT ACT_SITE 81 81
 CC FT ACT_SITE 194 194
 CC FT ACT_SITE 196 196
 CC SEQUENCE 553 AA; 60359 MW; 8D48C524F3058FC2 CRC64;
 CC
 CC Query Match 67.3%; Score 33; DB 1; Length 553;
 CC Best Local Similarity 87.5%; Pred. No. 31;
 CC Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC 2 FIGAINTD 9
 CC 179 FIGAIARD 186
 CC
 CC RESULT 14
 CC ID CHCC_ANTPO STANDARD: PRT: 167 AA.
 CC AC P08930;
 CC DT 01-NOV-1988 (Rel. 09; Created)
 CC DT 01-NOV-1997 (Rel. 35; Last sequence update)
 CC DT 01-NOV-1997 (Rel. 35; Last annotation update)
 CC DE CHORION CLASS CB PROTEIN PC404 (FRAGMENT).
 CC OS Antherea polyphemus (Polyphemus moth).
 CC CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 CC OC Bombycoidea; Saturniidae; Saturniinae; Antherea.
 CC OX NCBI_TaxID=7120;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=83195030; PubMed=6573656;
 CC RA Regier J.C., Kafatos F.C., Hamodrakas S.J.;
 CC RT *Silkmoth chorion multigene families constitute a superfamily:
 CC comparison of C and B family sequences.*
 CC RL Proc. Natl. Acad. Sci. U.S.A. 80:1043-1047(1983).
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
 CC SILK MOTH.
 CC -1- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH
 CC BELONG CLASSES B, CB AND HCB.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC
 CC EMBL: J01160; AAA27787.1; -
 CC DR EMBL: V00079; CAA23421.1; ALT TERM.
 CC DR InterPro: IPR002635; -
 CC DR Pfam: PF01723; Chorion; 1.
 CC KW Eggshell; Chorion; Repeat; Multigene family.
 CC FT DOMAIN 1 1
 CC FT ACT_SITE 55 55
 CC FT ACT_SITE 126 126
 CC FT ACT_SITE 127 >167
 CC FT NON_TER 167 167
 CC SEQUENCE 167 AA; 16202 MW; 428F08E9C3757382 CRC64;
 CC -----

Query Match 65.3%; Score 32; DB 1; Length 167;
 Best Local Similarity 62.5%; Pred. No. 15;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGA1ATD 9
 1:111111
 Db 94 FLGA1ATVD 101

RESULT 15
 MSHR_BOVIN STANDARD; PRT: 317 AA.
 AC P47798; Q28025;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE MELANOCYTE STIMULATING HORMONE RECEPTOR (MSH-R) (MELANOTROPIN RECEPTOR) (MELANOCORTIN-1 RECEPTOR) (MCL-R) (BDF3).
 GN MCIR OR MSHR.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94307438; PubMed=8034052;
 RA Vanetti M., Schoenrock C., Meyerhof W., Hoeltt V.;
 RT "Molecular cloning of a bovine MSH receptor which is highly expressed in the testis.";
 RT FEBS Lett. 348:268-272(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96325106; PubMed=8661706;
 RA Joerg H., Fries H.R., Meljerk E., Stranzinger G.F.;
 RT "Red coat color in Holstein cattle is associated with a deletion in the MSHR gene.";
 RT Mamm. Genome 7:317-318(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX Klungland H., Reed K.H., Vage D.L.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RT "A new melanocortin 1-receptor allele is coupled with Agouti coat colour pattern in Holstein cattle.";
 RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RECEPTOR FOR MSH (ALPHA, BETA) AND ACTH. DOES NOT SEEM TO BE ACTIVE WITH GAMMA-MSH. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLATE CYCLASE. COULD BE INVOLVED IN SPERMATOGENESIS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE TESTIS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CE -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: S71017; AAB31361.1; -;
 DR EMBL: 039469; AAC48590.2; -;
 DR EMBL: Y13957; CAA74291.1; -;
 DR EMBL: Y19103; CAB64818.1; -;
 DR GCRDB: GCR_1320; -;
 DR GCRDB: GCR_1337; -;
 DR GCRDB: GCR_2481; -;

DR InterPro: IPR000276; -;
 DR InterPro: IPR000761; -;
 DR InterPro: IPR001671; -;
 DR Pfam: PF00001; 7tm_1; 1.
 DR PRINTS: PRO0237; GPCRHDOPSN.
 DR PRINTS: PRO0534; MCRFAMILY.
 DR PRINTS: PRO0536; MELANOCYTESHR.
 DR PROSITE: PS00237; G-PROTEIN_RECIP_F1_1; 1.
 DR PROSITE: PS50262; G-PROTEIN_RECIP_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Phosphorylation; Lipoprotein; Palmitate.
 FT DOMAIN 1 37
 FT TRANSMEM 38 63
 FT DOMAIN 64 72
 FT TRANSMEM 73 93
 FT DOMAIN 94 118
 FT TRANSMEM 119 140
 FT DOMAIN 141 163
 FT TRANSMEM 164 183
 FT DOMAIN 184 191
 FT TRANSMEM 192 211
 FT DOMAIN 212 240
 FT TRANSMEM 241 266
 FT DOMAIN 267 279
 FT TRANSMEM 280 300
 FT DOMAIN 301 317
 FT CARBOHYD 29 29
 FT LIPID 315 315
 FT CONFLICT 40 41
 FT CONFLICT 99 99
 SQ SEQUENCE 317 AA; 34916 MW; 123692614F7FE168 CRC64;

Query Match 65.3%; Score 32; DB 1; Length 317;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGA1ATD 9
 1:1111111
 Db 134 FLGA1ATVD 141

Search completed: September 28, 2001, 12:39:12
 Job time: 787 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2001, 12:27:15 ; Search time 88.43 Seconds
(without alignments)
14.962 Million cell updates/sec

File: US-09-551-645-1
Perfect score: 49
Sequence: 1 TFFGATATDT 10

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

- Database :
- 1: SP-archaea:*
 - 2: SP-bacteria:*
 - 3: SP-fungi:*
 - 4: SP-human:*
 - 5: SP-invertebrate:*
 - 6: SP-mammal:*
 - 7: SP-mhc:*
 - 8: SP-organellar:*
 - 9: SP-phage:*
 - 10: SP-plant:*
 - 11: SP-rodent:*
 - 12: SP-unclassified:*
 - 13: SP-vertebrate:*
 - 14: SP-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	98.0	1603	2	09KKA6
2	48	98.0	1643	2	09FOP7
3	48	98.0	1643	2	09FOP6
4	40	81.6	520	2	007002
5	40	81.6	1616	2	09KKA9
6	39	79.6	1303	2	09XC45
7	39	79.6	1613	2	09KKB2
8	39	79.6	1614	2	09KKB6
9	39	79.6	1615	2	09KKA4
10	39	79.6	1615	2	09FOP9
11	39	79.6	1616	2	09KKB8
12	39	79.6	1616	2	09KKB3
13	39	79.6	1616	2	09KKB0
14	39	79.6	1616	2	09KKA7
15	39	79.6	1616	2	09KKA5
16	39	79.6	1616	2	09KKA2
17	39	79.6	1616	2	09KKA1
18	39	79.6	1616	2	09KKA0
19	39	79.6	1617	2	09KKB9

20	39	79.6	1617	2	09KKA3	09KKA3 rickettsia
21	39	79.6	1617	2	09KX98	09KX98 rickettsia
22	39	79.6	1618	2	09KKB4	09KKB4 rickettsia
23	39	79.6	1619	2	09KKB7	09KKB7 rickettsia
24	37	75.5	344	2	032258	032258 bacillus su
25	36	73.5	572	10	09FRU2	09FRU2 glycine max
26	36	73.5	1615	2	09KKA8	09KKA8 rickettsia
27	35	71.4	402	1	09HKW5	09HKW5 thermoplas
28	35	71.4	430	10	09SHC7	09shc7 arabidopsis
29	35	71.4	1304	2	09XC46	09XC46 rickettsia
30	35	71.4	1604	2	09KX99	09KX99 rickettsia
31	35	71.4	1618	2	09KKB1	09KKB1 rickettsia
32	35	71.4	1620	2	09KKB5	09KKB5 rickettsia
33	34	69.4	287	2	09RCX1	09RCX1 streptomyce
34	34	69.4	399	2	053522	053522 mycobacteri
35	34	69.4	468	2	09Z806	09Z806 chlamydia p
36	33	67.3	299	5	09XXA2	09XXA2 caenorhabdi
37	33	67.3	302	2	09X4M5	09X4M5 lactobacill
38	33	67.3	314	5	023017	023017 caenorhabdi
39	33	67.3	342	5	09XNV2	09XNV2 dirosophila
40	33	67.3	385	2	09HXM2	09HXM2 pseudomonas
41	33	67.3	389	14	09WT45	09WT45 human herpe
42	33	67.3	538	2	09L6T6	09L6T6 escherichia
43	33	67.3	555	2	033064	033064 mycobacteri
44	33	67.3	615	5	09VDF1	09VDF1 dirosophila
45	33	67.3	776	8	09XOV2	09XOV2 heterocapsa

ALIGNMENTS

RESULT 1

ID 09KKA6 PRELIMINARY; PRT; 1603 AA.

AC 09KKA6; 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE OMPB (FRAGMENT).

GN OMPB.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiidae; Rickettsia.

OX NCBI_TaxID=782;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BREINL;

RX PubMed=10939649;

RA Roux V., Raoult D.;

RT "Phylogenetic analysis of members of the genus Rickettsia using the

RT gene coding the outer-membrane protein ompB (ompB).";

RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).

DR EMBL; AF123718; AAF34121.1; -

FT SEQUENCE 1603 1603

SO SEQUENCE 1603 AA; 165704 MW; 6D513B6DCDDF6982 CRC64;

Query Match 98.0%; Score 48; DB 2; Length 1603;

Best Local Similarity 90.0%; Pred. No. 0.7;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TFFGATATDT 10

DB 1263 TFFGATATDT 1272

RESULT 2

ID 09FOP7 PRELIMINARY; PRT; 1643 AA.

AC 09FOP7; 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE OUTER MEMBRANE PROTEIN B.
GN OMPB.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FLORIDA;
RA Moron C.G., Bouyer D.H., Yu X.-J., Foll L.D., Crocquet-Valdes P.,
RA Walker D.H.;
RT "Phylogenetic analysis of the rmp B genes of Rickettsia felis and
RT Rickettsia prowazekii European human and North American flying
RT squirrel strains";
RL submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF211820; AAC43541.1; -
SO SEQUENCE 1643 AA; 169818 MW; 68FFB8CE87E496B5 CRC64;

Qy 1 TFICAIATDT 10
Db 1296 TFGAIAATDT 1305

RESULT 3
Q9F0P6 PRELIMINARY; PRT; 1643 AA.
AC Q9F0P6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OUTER MEMBRANE PROTEIN B.
GN OMPB.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VIRGINIA;
RA Moron C.G., Bouyer D.H., Yu X.-J., Foll L.D., Crocquet-Valdes P.,
RA Walker D.H.;
RT "Phylogenetic analysis of the rmp B genes of Rickettsia felis and
RT Rickettsia prowazekii European human and North American flying
RT squirrel strains";
RL submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF211821; AAC43542.1; -
SO SEQUENCE 1643 AA; 169767 MW; D22EF05D9D801AAE CRC64;

Query Match 98.0%; Score 48; DB 2; Length 1643;
Best Local Similarity 90.0%; Pred. No. 0.72;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 4
ID 007002 PRELIMINARY; PRT; 520 AA.
AC 007002;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHEICAL 57.0 KDA PROTEIN.
GN YVEA.
RT Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Denizot F.C.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384337;
RA Kunst F., Ogasawara N., Moszer J., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano N.J., Carter R.A.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
RA Ertian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Melado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogawa K., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadle Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solito B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takamori K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vaasacoti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Yasumoto K., Yata K.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z94043; CAB08050.1; -
DR EMBL: Z99121; CAB15452.1; -
DR InterPro: IPR001899; -
DR InterPro: IPR002027; -
DR InterPro: IPR002293; -
DR Pfam: PF00324; aa_permeases; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR Hypothetical protein.
SO SEQUENCE 520 AA; 56988 MW; 0583CB56FE6D37D9 CRC64;

Query Match 81.6%; Score 40; DB 2; Length 520;
Best Local Similarity 80.0%; Pred. No. 8.2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TFICAIATDT 10
Db 253 TFGAIAPIET 262

RESULT 5
Q9KKA9 PRELIMINARY; PRT; 1616 AA.
AC Q9KKA9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

```

DE 01-MAR-2001 (TREMblrel. 16, last annotation update)
GN OMPB (FRAGMENT).
OS Rickettsia mongolotimonae.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=45261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MARSEILLE 1;
RX PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RL gene coding the outer-membrane protein ompB (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR EMBL; AF123715; AAF34118.1; -.
FT NON_TER 1616
SQ SEQUENCE 1616 AA; 164312 MW; 9ACFBEC7062AD47 CRC64;

Query Match
Best Local Similarity 81.6%; Score 40; DB 2; Length 1616;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 FIGA1ATDT 10
DB 1269 SFVGAIVTDT 1278

RESULT 6
Q9KX45 PRELIMINARY; PRT; 1303 AA.
AC Q9KX45;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE OUTER MEMBRANE PROTEIN B (FRAGMENT).
GN OMPB.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MALISH 7;
RA Stenos J., Walker D.;
RT "The rickettsial outer membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF149110; AAD39533.1; -.
DR InterPro; IPR001639; -.
DR PRINTS; PRO0810; BCTERIALGSPC.
FT NON_TER 1
SQ SEQUENCE 1303 AA; 132895 MW; 30F092863A6CC35D CRC64;

Query Match
Best Local Similarity 79.6%; Score 39; DB 2; Length 1303;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGA1ATDT 10
DB 951 FVGAIVTDT 959

RESULT 7
Q9KKB2 PRELIMINARY; PRT; 1613 AA.
AC Q9KKB2;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, last annotation update)
DE OMPB (FRAGMENT).

```

```

GN OMPB.
OS Israeli tick typhus rickettsia.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=45258;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISTT CDC1;
RX PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RL gene coding the outer-membrane protein ompB (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR EMBL; AF123712; AAF34115.1; -.
FT NON_TER 1613
SQ SEQUENCE 1613 AA; 164218 MW; A1F278D7F80FD829 CRC64;

Query Match
Best Local Similarity 79.6%; Score 39; DB 2; Length 1613;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGA1ATDT 10
DB 1270 FVGAIVTDT 1278

RESULT 8
Q9KKB6 PRELIMINARY; PRT; 1614 AA.
AC Q9KKB6;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE OMPB (FRAGMENT).
GN OMPB.
OS Astrakhan rickettsia.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=140892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A-167;
RX PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein ompB (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR EMBL; AF123708; AAF34111.1; -.
FT NON_TER 1614
SQ SEQUENCE 1614 AA; 164176 MW; 3930D285771CDA98 CRC64;

Query Match
Best Local Similarity 79.6%; Score 39; DB 2; Length 1614;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGA1ATDT 10
DB 1271 FVGAIVTDT 1279

RESULT 9
Q9KKA4 PRELIMINARY; PRT; 1615 AA.
AC Q9KKA4;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, last annotation update)
DE OMPB (FRAGMENT).
OS Rickettsia sp. S.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

```

OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=45263;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN STRAIN S;
RX PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein ompB (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR EMBL: AF123720; AAF34123.1; -
FT NON_TER 1615 1615
SQ SEQUENCE 1615 AA; 164362 MW; 239387B0B84E267B CRC64;

Query Match 79.6%; Score 39; DB 2; Length 1615;
Best Local Similarity 77.8%; Pred. No. 45;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGAATDT 10
DB 1269 FVGATVDT 1277

RESULT 10
O9F0P9 PRELIMINARY; PRT; 1615 AA.
AC O9F0P9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OUTER MEMBRANE PROTEIN OMPB (FRAGMENT).
GN OMPB.
OS Rickettsia sp. California 2.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=147259;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CALIFORNIA 2;
RX Raoult D.;
RT "A new SFG rickettsia isolated from fleas.";
RT Submitted (Jan-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CALIFORNIA 2;
RX Roux V., Raoult D.;
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF210695; AAG48556.1; -
FT NON_TER 1615 1615
SQ SEQUENCE 1615 AA; 163691 MW; 7A69C95827E20DA8 CRC64;

Query Match 79.6%; Score 39; DB 2; Length 1615;
Best Local Similarity 77.8%; Pred. No. 45;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGAATDT 10
DB 1270 FVGATVDT 1278

RESULT 11
O9KKB8 PRELIMINARY; PRT; 1616 AA.
AC O9KKB8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OMPB (FRAGMENT).
GN OMPB.
OS Rickettsia africana.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35788;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ESF-5;
RX PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein ompB (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR EMBL: AF123706; AAF34109.1; -
FT NON_TER 1616 1616
SQ SEQUENCE 1616 AA; 164248 MW; E4B06C912859D5A6 CRC64;

Query Match 79.6%; Score 39; DB 2; Length 1616;
Best Local Similarity 77.8%; Pred. No. 45;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGAATDT 10
DB 1270 FVGATVDT 1278

RESULT 12
O9KKB3 PRELIMINARY; PRT; 1616 AA.
AC O9KKB3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OMPB (FRAGMENT).
GN OMPB.
OS Rickettsia honei.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=37816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB;
RX PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein ompB (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR EMBL: AF123711; AAF34114.1; -
FT NON_TER 1616 1616
SQ SEQUENCE 1616 AA; 164500 MW; 8053C358A44A72D9 CRC64;

Query Match 79.6%; Score 39; DB 2; Length 1616;
Best Local Similarity 77.8%; Pred. No. 45;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGAATDT 10
DB 1270 FVGATVDT 1278

RESULT 13
O9KKB0 PRELIMINARY; PRT; 1616 AA.
AC O9KKB0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OMPB (FRAGMENT).
GN OMPB.
OS Rickettsia massilliae.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35791;

RN (1)
 RC SEQUENCE FROM N.A.
 RX STRAIN=MT01;
 RA PubMed-10939649;
 RT "Phylogenetic analysis of members of the genus *Rickettsia* using the
 gene coding the outer-membrane protein ompB (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 DR EMBL: AF123714; AAF34117.1; -;
 FT NON_TER 1616
 SQ SEQUENCE 1616 AA; 162964 MW; 9CF6EA2BA1B7821D CRC64;

Query Match 79.6%; Score 39; DB 2; Length 1616;
 Best Local Similarity 77.8%; Pred. No. 45;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 OY 2 FIGA1ATDT 10
 Db 1270 FVGA1VTD 1278

RESULT 14
 O9KKA7 PRELIMINARY; PRT; 1616 AA.
 AC O9KKA7;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
 DE OMPB (FRAGMENT).
 GN OMPB.
 OS *Rickettsia parkeri*.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=35792;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MACULATUM 20;
 RX PubMed-10939649;
 RA Roux V., Raoult D.;
 RT "Phylogenetic analysis of members of the genus *Rickettsia* using the
 gene coding the outer-membrane protein ompB (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 DR EMBL: AF123717; AAF34120.1; -;
 FT NON_TER 1616
 SQ SEQUENCE 1616 AA; 164254 MW; 3D270DF9D5C16208 CRC64;

Query Match 79.6%; Score 39; DB 2; Length 1616;
 Best Local Similarity 77.8%; Pred. No. 45;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGA1ATDT 10
 Db 1270 FVGA1VTD 1278

RESULT 15
 O9KKA5 PRELIMINARY; PRT; 1616 AA.
 AC O9KKA5;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
 DE OMPB (FRAGMENT).
 GN OMPB.
 OS *Rickettsia rhipicephali*.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=33992;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=3-7-6;

RX PubMed-10939649;
 RA Roux V., Raoult D.;
 RT "Phylogenetic analysis of members of the genus *Rickettsia* using the
 gene coding the outer-membrane protein ompB (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 DR EMBL: AF123719; AAF34122.1; -;
 FT NON_TER 1616
 SQ SEQUENCE 1616 AA; 163270 MW; 42E0489239F15EB4 CRC64;

Query Match 79.6%; Score 39; DB 2; Length 1616;
 Best Local Similarity 77.8%; Pred. No. 45;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 OY 2 FIGA1ATDT 10
 Db 1270 FVGA1VTD 1278

Search completed: September 28, 2001, 12:40:52
 Job time: 817 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 2001, 10:00:15 ; Search time 1261.48 Seconds
(without alignments)
367.848 Million cell updates/sec

Title: US-09-551-645-2
Perfect score: 30
Sequence: 1 acattataggagctatgtctacagataca 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_cm:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_cm:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_stal:*
54: gb_sts1:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_vl2:*
60: gb_hg1:*
61: gb_hg2:*
62: gb_hg3:*
63: gb_hg4:*
64: gb_hg5:*
65: gb_hg6:*
66: gb_hg7:*
67: gb_hg8:*
68: gb_hg9:*
69: gb_hg10:*
70: gb_hg11:*
71: gb_hg12:*
72: gb_hg13:*
73: gb_hg14:*
74: gb_hg15:*
75: gb_hg16:*
76: gb_hg17:*
77: gb_hg18:*
78: gb_hg19:*
79: gb_hg20:*
80: gb_hg21:*
81: gb_hg22:*
82: gb_hg23:*
83: gb_hg24:*
84: gb_hg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_rol2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	5258	3	RIRCSUPT
2	28.4	94.7	4851	1	AF123718
3	28.4	94.7	5015	2	AF161079
4	28.4	94.7	5015	2	AF211820
5	28.4	94.7	5015	2	AF211821
6	28.4	94.7	5131	3	RIRSPAP
7	28.4	94.7	5319	9	AR019483
8	23	76.7	2794	7	AF284000

C	9	21.6	72.0	101384	90	AL139014	Human DNA
	10	21.6	72.0	164247	79	AL158082	Homo sapi
	11	21.6	72.0	171713	80	AL156152	Homo sapi
	12	21.4	71.3	2785	7	AF283999	Dasyatis
	13	21.1	70.0	91859	90	AL158133	Human DNA
C	14	21.1	70.0	220216	88	AC021019	Homo sapi
	15	20.8	69.3	2773	7	AF284019	Dugong du
	16	20.8	69.3	2788	7	AF284030	Oryzopsis
	17	20.8	69.3	2794	7	AF284020	Trilichocheu
	18	20.6	68.7	2830	2	BFXYNB	B. fibrioloso
	19	20.6	68.7	4629	5	AF273674	Plasmodiu
C	20	20.6	68.7	4846	2	AF210695	Rickettsi
	21	20.6	68.7	4900	1	AF123707	Rickettsi
	22	20.6	68.7	167629	69	AC024974	Homo sapi
	23	20.6	68.7	231916	79	AL158037	Homo sapi
	24	20.4	68.0	31754	11	AF009630	Bacteriop
	25	20.4	68.0	37667	11	AF304433	Bacteriop
	26	20.4	68.0	74999	87	AC008960	Homo sapi
	27	20.4	68.0	139874	93	HSJ91818	Human DNA
	28	20.4	68.0	139012	61	AC010503	Homo sapi
	29	20.4	68.0	144014	62	AC011685	Homo sapi
	30	20.4	68.0	146432	65	AC019022	Homo sapi
	31	20.4	68.0	152195	64	AC016331	Homo sapi
	32	20.4	68.0	155025	88	AC034246	Homo sapi
	33	20.4	68.0	155805	83	AF001987	Homo sapi
	34	20.4	68.0	160889	90	AL156335	Human DNA
	35	20.4	68.0	161239	64	AC016580	Homo sapi
	36	20.4	68.0	162868	74	AC072027	Homo sapi
	37	20.4	68.0	170088	69	AC025119	Homo sapi
	38	20.4	68.0	171274	70	AC026476	Homo sapi
	39	20.4	68.0	179060	75	AC074382	Homo sapi
	40	20.4	68.0	183936	94	AC007978	Mus muscu
	41	20.4	68.0	184559	73	AC069017	Mus muscu
	42	20.4	68.0	336619	74	AC073522	Homo sapi
	43	20.2	67.3	132124	60	AC004737	Homo sapi
	44	20.2	67.3	176601	69	AC026049	Homo sapi
C	45	20.2	67.3	197730	65	AC019094	Homo sapi

ALIGNMENTS

RESULT 1	RIRCSLPT	5258 bp	DNA	BC1	10-MAR-1997
LOCUS	Rickettsia typhi crystalline surface layer protein (slpr)				
DEFINITION	gene complete cds.				
ACCESSION	L04661	GI:152474			
VERSION	L04661.1	GI:152474			
KEYWORDS	Rickettsia typhi.				
SOURCE	Rickettsia typhi.				
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.				
REFERENCE	1 (bases 1 to 5258)				
AUTHORS	Hahn, W.-J., Kim, K.-K., Kim, I., and Chang, W.-H.				
TITLE	Cloning and sequence analysis of the gene encoding the crystalline surface layer protein of Rickettsia typhi				
JOURNAL	Gene 133 (1), 129-133 (1993)				
MEDLINE	94040787				
FEATURES	Location/Qualifiers				
source	1..5258				
	/organism="Rickettsia typhi"				
	/strain="Williamson"				
	/db_xref="taxon:785"				
	1..>5258				
gene	/gene="slpr"				
5'UTR	<1..143				
CDS	/gene="slpr"				
	144..5081				
	/codon_start=1				
	/transl_table=11				

Query Match	100.0%	Score 30;	DB 3;	Length 5258;
Best Local Similarity	100.0%	Pred. 0.03;		
Matches	30;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
Qy	1	acattatagagcctatgctacagatca	30	
Db	4029	ACATTATAGAGCCTATGCTACAGATCA	4058	

3'UTR

BASE COUNT 1757 a 755 c 1008 g 1738 t

ORIGIN

/product="crystalline surface layer protein"

/protein_id="AAB48987.1"

/db_xref="GI:1871223"

/translation="MAQPNFLKIIISAGLVSTATIVAGSGVAMGAVMQNRTTN
AAATVDGAGFDQTGAVNLVPVATNSVTANSNNAITFPNPNLNSLFDIANTLAV
TINENTLTGVTNTKOGNFENFTIGKSLITIGHEITAGQAATTSAQNVSKVA
GAINDNDLSGVSDIFETAPSVLEFLIPTDTEGALITGDAKIVNGAGLINTN
GPKVSPDKTITAGKITITIGNOGIMFTPTDANALNLDGGGNTINFGDGGKVL
VSKNGNATERNVTSGLGKDIYEFPTTAAAGKLINGAANNAVICTDAGRAAG
IVSDNGNAAITSGOVYAKDVIQSAAGQVFEHLVDVGLGKTFKTDASKVIT
ENASFGSTDEGNLAVQIVPNPKILTGNFEGDAKNNGNTAGVITFANPNQVLYNN
NIYVNTIKALEVEGAGIVOLSGIGALRGNAGSIFKLDGNTVFNPNQVLYNN
ALMAGSTOLDGSAITLTDGNGAVNALADITLNDASKITLTSAGNIIGAGANK
FOANGTIOITSTNNITLVPDPTDYTDQGVDAISLTNQTITNGSTIGTGANR
TLGRFNVSSTKITLMDGVAINELVMDNDSSVHLTHNTYLTITKANANGKITVAD
PIHNDTADLTADNTNGSAESINTEFTKANGOSILHDKGVNLVANNITTDANG
SIHRSRGTSIVSGTGGGGLKNNILNNGTIVKFLGDTFNGKIEGSLQIS
SNYITDIESADNTGLEFNTDPTITVTLNKGAFVGLKQVMSGPNLAVFNGIG
VAHIAVDSTSEFNASIGASIFLISGRPLDITKISVNGQVDFNAPLIVSGIDS
MINNGOYIGDKNITIALSLGSDNSITVNSNTVAGITPTKNGSTVTLSGGIPNPGT
IYGLGNGDKRLKQVFTFYDYNLGSITIRATNTINDVTLTGGIAGTDPDKITIG
SINGNAVKEVDRTFSHPSTSMIVSTKANGVITVYLGALVNGTSSDI PASVRETN
DSGVLOGNHSQNIIDEGTYNLITLNSDVLIGGTTAINEIDLTNNLLFANTSTW
GNMSTLSTTLVNSNGVNGOIVYIAEGAQVNAATTTGTTIKIODNANANFSGTOTYTLIO
AGAPNGTCLAPNEDVGNMTEFVKEVLEIROANOYVLTFRNDVNVVTVAVNSAIAN
APCHONAIATLESTDPRAVNNMLAARDSSVAFETAIATPDGAATVAVNLNDPOTK
QDLGNRIGALRYLSNETADVSGSEPGAASGDEAIDQVSYGMAPFNIAADPK
GGLAGYKRAKAGVAVGADTTLNMLMGAIGITRTIKHODVKKGKIDIKLSFSL
YGAOLVKNFPAQSAIFTLNKKVSKSQREFPDANGKMNQIAAGANDNTFGGKLMF
GYDNAQVGVLPFPMAGISYLSKSNENYKEGTVANKRKPSRIDILVGAKEVG
SAHNDIVYVPEIHSFYHVKNGLSKAOSMLDGGAPFISQDPRAKNSYNIGLSA
NIRSDAKMEIGTIGDFNNAKRYTHOGTLKRYNF"

/gene="slpr"

5082..>5258

Query Match	100.0%	Score 30;	DB 3;	Length 5258;
Best Local Similarity	100.0%	Pred. 0.03;		
Matches	30;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
Qy	1	acattatagagcctatgctacagatca	30	
Db	4029	ACATTATAGAGCCTATGCTACAGATCA	4058	

RESULT 2

AF123718 4851 bp DNA BCT 09-AUG-2000

LOCUS Rickettsia prowazekii ompb (ompb) gene, partial cds.

DEFINITION AF123718

ACCESSION AF123718

VERSION AF123718.1 GI:6969951

KEYWORDS Rickettsia prowazekii.

SOURCE Rickettsia prowazekii.

ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.

REFERENCE 1 (bases 1 to 4851)

AUTHORS Roux, V., and Raoult, D.

TITLE Phylogenetic analysis of members of the genus Rickettsia using the gene encoding the outer-membrane protein ompb

JOURNAL Int. J. Syst. Evol. Microbiol. 50 Pt 4, 1449-1455 (2000)

MEDLINE 20393643

REFERENCE 2 (bases 1 to 4851)

AUTHORS Roux, V., and Raoult, D.

TITLE Direct Submission

JOURNAL Submitted (27-JUN-1999) Unite des Rickettsies, Faculte de Medecine, 27 Boulevard Jean Moulin, Marseille 13385, France

FEATURES

source

1..4851

/organism="Rickettsia prowazekii"

/strain="Breinl"

JOURNAL	squirrel strains		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 5015)		
TITLE	Moron,C.G., Bouyer,D.H. and Walker,D.H.		
JOURNAL	Submitted (06-DEC-1999) Pathology, University of Texas Medical Branch at Galveston, 300 University Blvd., Galveston, TX 77555, USA		
FEATURES	Location/Qualifiers		
source	1. .5015		
	/organism="Rickettsia prowazekii"		
	/strain="Florida"		
gene	/db_xref="taxon:782"		
	84. .5015		
	/gene="ompB"		
	84. .5015		
CDS	/gene="ompB"		
	/codon_start=1		
	/transl_table=11		
	/product="outer membrane protein B"		
	/protein_id="Aa63541.1"		
	/db_xref="GI:12003368"		
	/translation="MAQKPNFLKKIISAGLVSTASTATIVAGFSGVANGAAMQVNRNTN		
	AAATPFGIGEDQAAGANIPVAPNSVITANANNITFNTFNGHLSLEFDLANDLAVT		
	INDDTGTGFTNTAQAOKAFENFVPAACKLINTGCGITVGEASNTINAONALTKAHGG		
	AAINANDSLGSGITTPAASVLEFNINPTQAPLITLQANSKIVGNGSLITITNG		
	FIVSDVTFEAGIKTINIDDCQGLMFNSTPDANLNLQVGNITFNGIDGTGKLVY		
	SKNGAATFNTGTLGNLKGIELNTAAVAGKLIISOGANAVITGDNAGRAAGFI		
	VSDVNGAATISGQVYAKNMVIGSANAGQVTEFHIYDVGLGTTNFKTADSKYITE		
	NSNGSTNFGNDLOIVVPTDKILKNGFIDGVKNKGNTAGVITFNGALVASTDPN		
	IATVNTAIAEAGAGVLESGIHAEIRLNGSGIFKLADGTVNGPNOVALANNA		
	LAAGSITLDSASITTGDIINGGVNAAQLHTLNDASKITALDQANTIGAVGAHIF		
	QANGGTRKLTNTQNNIVNFDLITDKAGVNDASSLTNNOTLTINSIGTVANRT		
	LAOLNIGSSKTIILNAGVAINELVIENTNGVOLNHNRYLTKTINANOQOIIYAAP		
	LNTWTTLADGTNLGSAENPLSTIFATKANADSILWVGKVNLYANNITINDANVS		
	LHFRSGGTSIVSGVGGQGHKLNNLLDNGTVKFLGDTTFNGGTRIEGSILOISN		
	NNTDYHESADNCGTLEFVNTPDPTVTLNKGAVFGLKQVLIISPGNIYFNEIGNG		
	IVHGIAANSISFENASLGTSILFPLSGTPDLVLPKTSVNGCTVNDPNAPIVYSGIDS		
	MINNGQIIGDKKNITIALSLSDNSITVANNTLSGIRPTKNNQSTVLSGMPNPGT		
	IYGLGELGSPKRLQVFTTDDYNLGSIIANNVTIINDVTLTTGGIAGTDEDAITLG		
	SVNGANAVREVDSTFSDPSRMIVATQANKGVYLGNALVNSIGSLDTPVASFETGN		
	DSGGLQGNITYSQNIIDGTYNLTILNSNVLIGGTTAINGEIDLTLNNLIFANGSTY		
	GDNISITTLNVSNGNIGOVYIADQAVNATTTGTTIKIODNANANPSCGOAVTLIO		
	GGARFNGTILGAPNFAVGSNI FVKEELIRSDNDVYLTRINDVNLVTTAVGNSALIN		
	APGVSONISRCLESTNTAAVNNMLLARDPSDVATFEVAILATDYSAAVTVALNDTKT		
	QDLNSRLGTLRLYLSNAETSDVAGSATGAVSSGDEAVSYGVMAKPEYINAEODKKG		
	IAGYKAKTGTGVVGLDITLASDNLMIGAIGITKTDIHDQYKKGDKTDINLSFSLG		
	SOOLVKNFPAQNAIFTLNKKSKSORFPFESNGMSKOIAGAENDMTFEGNLIIFY		
	DYNAKPNVLVTPMAGLSYLKSSNENYKETGTVANKNRINSKESRVDLIVGAVAGST		
	VNITDIYIPEIHSFVYHVKNGKLSNSQSLDGTAPFISOPDRTAKTSYINIGLSANI		
	KSDAKMEYGIQYDENSASKYTAHQGLTKVRNF"		
BASE COUNT	1697 a	731 c	950 g 1637 t
ORIGIN			
Query Match	94.7%;	Score 28.4;	DB 2; Length 5015;
Best Local Similarity	96.7%;	Pred. No. 0.14;	
Matches	29; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
OY	1	acattatagagcctatgtcacagataca	30
Db	3969	ACATTGTGAGGAGCTATTGCTACAGATACA	3998
RESULT	5		
LOCUS	AF211821	5015 bp	DNA
DEFINITION	Rickettsia prowazekii strain Virginia outer membrane protein B		
ACCESSION	AF211821		
VERSION	AF211821.1	GI:12003369	
KEYWORDS	Rickettsia prowazekii.		
SOURCE	Rickettsia prowazekii		
ORGANISM			

JOURNAL	Bacteria: Proteobacteria: alpha subdivision: Rickettsiales;		
REFERENCE	Rickettsiaceae; Rickettsiae; Rickettsia: typhus group.		
AUTHORS	1 (bases 1 to 5015)		
TITLE	Moron,C.G., Bouyer,D.H., Yu,X.-J., Foll,L.D., Crocquet-Valdes,P. and Walker,D.H.		
JOURNAL	Phylogenetic analysis of the omp B genes of Rickettsia felis and Rickettsia prowazekii European human and North American flying squirrel strains		
FEATURES	Location/Qualifiers		
source	1. .5015		
	/organism="Rickettsia prowazekii"		
	/strain="Virginia"		
gene	/db_xref="taxon:782"		
	84. .5015		
	/gene="ompB"		
	84. .5015		
CDS	/gene="ompB"		
	/codon_start=1		
	/transl_table=11		
	/product="outer membrane protein B"		
	/protein_id="Aa63542.1"		
	/db_xref="GI:12003370"		
	/translation="MAQKPNFLKKIISAGLVSTASTATIVAGFSGVANGAAMQVNRNTN		
	AAATPFGIGEDQAAGANIPVAPNSVITANANNITFNTFNGHLSLEFDLANDLAVT		
	INDDTGTGFTNTAQAOKAFENFVPAACKLINTGCGITVGEASNTINAONALTKAHGG		
	AAINANDSLGSGITTPAASVLEFNINPTQAPLITLQANSKIVGNGSLITITNG		
	FIVSDVTFEAGIKTINIDDCQGLMFNSTPDANLNLQVGNITFNGIDGTGKLVY		
	SKNGAATFNTGTLGNLKGIELNTAAVAGKLIISOGANAVITGDNAGRAAGFI		
	VSDVNGAATISGQVYAKNMVIGSANAGQVTEFHIYDVGLGTTNFKTADSKYITE		
	NSNGSTNFGNDLOIVVPTDKILKNGFIDGVKNKGNTAGVITFNGALVASTDPN		
	IATVNTAIAEAGAGVLESGIHAEIRLNGSGIFKLADGTVNGPNOVALANNA		
	LAAGSITLDSASITTGDIINGGVNAAQLHTLNDASKITALDQANTIGAVGAHIF		
	QANGGTRKLTNTQNNIVNFDLITDKAGVNDASSLTNNOTLTINSIGTVANRT		
	LAOLNIGSSKTIILNAGVAINELVIENTNGVOLNHNRYLTKTINANOQOIIYAAP		
	LNTWTTLADGTNLGSAENPLSTIFATKANADSILWVGKVNLYANNITINDANVS		
	LHFRSGGTSIVSGVGGQGHKLNNLLDNGTVKFLGDTTFNGGTRIEGSILOISN		
	NNTDYHESADNCGTLEFVNTPDPTVTLNKGAVFGLKQVLIISPGNIYFNEIGNG		
	IVHGIAANSISFENASLGTSILFPLSGTPDLVLPKTSVNGCTVNDPNAPIVYSGIDS		
	MINNGQIIGDKKNITIALSLSDNSITVANNTLSGIRPTKNNQSTVLSGMPNPGT		
	IYGLGELGSPKRLQVFTTDDYNLGSIIANNVTIINDVTLTTGGIAGTDEDAITLG		
	SVNGANAVREVDSTFSDPSRMIVATQANKGVYLGNALVNSIGSLDTPVASFETGN		
	DSGGLQGNITYSQNIIDGTYNLTILNSNVLIGGTTAINGEIDLTLNNLIFANGSTY		
	GDNISITTLNVSNGNIGOVYIADQAVNATTTGTTIKIODNANANPSCGOAVTLIO		
	GGARFNGTILGAPNFAVGSNI FVKEELIRSDNDVYLTRINDVNLVTTAVGNSALIN		
	APGVSONISRCLESTNTAAVNNMLLARDPSDVATFEVAILATDYSAAVTVALNDTKT		
	QDLNSRLGTLRLYLSNAETSDVAGSATGAVSSGDEAVSYGVMAKPEYINAEODKKG		
	IAGYKAKTGTGVVGLDITLASDNLMIGAIGITKTDIHDQYKKGDKTDINLSFSLG		
	SOOLVKNFPAQNAIFTLNKKSKSORFPFESNGMSKOIAGAENDMTFEGNLIIFY		
	DYNAKPNVLVTPMAGLSYLKSSNENYKETGTVANKNRINSKESRVDLIVGAVAGST		
	VNITDIYIPEIHSFVYHVKNGKLSNSQSLDGTAPFISOPDRTAKTSYINIGLSANI		
	KSDAKMEYGIQYDENSASKYTAHQGLTKVRNF"		
BASE COUNT	1694 a	735 c	951 g 1635 t
ORIGIN			
Query Match	94.7%;	Score 28.4;	DB 2; Length 5015;
Best Local Similarity	96.7%;	Pred. No. 0.14;	
Matches	29; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
OY	1	acattatagagcctatgtcacagataca	30
Db	3969	ACATTGTGAGGAGCTATTGCTACAGATACA	3998
RESULT	6		
LOCUS	AF211821	5131 bp	DNA
DEFINITION	Rickettsia prowazekii		
ACCESSION	AF211821		
VERSION	AF211821.1	GI:12003369	
KEYWORDS	Rickettsia prowazekii		
SOURCE	Rickettsia prowazekii		
ORGANISM			

[illegible]

IIQHSILJNKKIRKRRRTTSCLEHEDFTKAYDLTVQKTPKINOGIDQMEQNOVANS
ANGSHENETKGDVQKKEENANPIESLESKEAFRTKEPSSISINMELNITNSKAS
KXRLRMSSTRIHAIELVGNRPSPKTPLEQIOJSCSTEEIEKINSQKPIRHR
MLQMEKENTGAKNNKPNQEOISEHSHADPELKLITVTEFLPEKSNQKDIQEV
NSSLQEVANLETTIOVSDTRNPEIDTSEGSLQORAEISISVVDPTGYQGD
SVSLLEDITGKAKTALNOPMSOYAAKNAKNTESLHGCAKTRNDTEDEKPLAEVH
TOETSVEMSESELDOTLOLWTFKISKQSFALSPNEBCATVCAHSRELGKSPKYI
FECROKEENQKESKIKHVAHTTGAFFVQCKDPGYACSIQEVSRICQSSQF
RGNESELITANEHEISQNPOMPSLIRSSVTKCKENLSEERFELTISLERTVGN
ENIVOSTVSTISHNNIRENAFKSSSINEVSSGENIOALEGRNAPKLNMLRLG
LMQPEYKOSLPITTSCKHPETKROGENEALQAVDPFPHLISDLNLEPMGNSHVQ
ICSETPDDLEDDDEIKENNSFAEBGIERSAVSKSQKKEFRSPPLHTLSLAQCH
QRRARKIDSSSEEVSE"

BASE COUNT 1052 a 495 c 575 g 671 t 1 others

ORIGIN

Query Match 76.7% Score 23; DB 7; Length 2794;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttataggagctattgctacaga 26
|||||
Db 626 TTATAGAGCTATTGCTACAGA 648

RESULT 9
AL139014/c
LOCUS
DEFINITION Human DNA sequence from clone RP4-586015 on chromosome 1p21.2-22.2
Contains part of a gene for a hypothetical 22.1 kDa protein, ESTs,
STSs and GSSs, complete sequence.
ACCESSION AL139014
VERSION AL139014.6 GI:9187181
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 101384)
Lloyd, D.
Direct Submission
Submitted (04-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jul 14, 2000 this sequence version replaced gi:8250669.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone comp1 of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP4-586015 is from
the library RPI-4 constructed by the group of Pieter de Jong. For
further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP4-586015 The true

left end of clone RPI1-14019 is at 31720 in this sequence. The true
right end of clone RPI1-57H12 is at 50350 in this sequence. The
true right end of clone RPI1-335D10 is at 62933 in this sequence.
location/Qualifiers
1..101384
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p21.2-22.2"
/clone="RP4-586015"
/clone_1bp="RPI-4"
1140..1549
/note="match: STS: Em:HSPB32C12"
3609..4115
/note="LTR22 repeat: matches 1..580 of consensus"
join(<6174..6678,8719..9293)
/gene="dJ586015.1"
match: ESTs: Em:AI035684 Em:AI591890 Em:AI662458
Em:AA177712 Em:AM527581 Em:AM269377 Em:AA115559
Em:AA336662 Em:AI344603 Em:AA825527 Em:AI457123
Em:AA483584 Em:AI870117 Em:RI9361 Em:N80756 Em:AI499762
Em:AI566566 Em:R07610 Em:AM613686 Em:AA336253 Em:AA770533
Em:AI023354 Em:AA699431 Em:AI187083 Em:AI969585"
/evidence="not_experimental"
/product="dJ586015.1 (hypothetical 22.1 kDa protein)"
6174..9293
/gene="dJ586015.1"
<6174..6676
/gene="dJ586015.1"
/note="match: proteins: Tr:Q9Y3V2"
/codon_start=3
/evidence="not_experimental"
/product="dJ586015.1 (hypothetical 22.1 kDa protein)"
/protein_id="CAC36064.1"
/db_xref="GI:13559035"
/translation="TDOCTVPRIRHTRKAGFMDADIPLEVHLPLVNPVSCPLGJINSE
QLTRAQCVTVKENILEQAESLSPPMYHELVTWQNLRIHLSPETGSSSECTPST
STTMDDLWTLLHLHDMRAKTKYKIVKAWASDLRLTGLRMFGKILLILQGDNRN
LKVPKS"
complement(6407..6852)
/note="match: STS: Em:HSPB52C06"
6849..7492
/gene="dJ586015.1"
/note="match: GSS: Em:AQ021020"
complement(8370..8651)
/note="match: GSS: Em:AQ087817"
9274..9279
/gene="dJ586015.1"
9293
/gene="dJ586015.1"
9984..10015
/note="16 copies 2 mer tt 84% conserved"
complement(10605..11104)
/note="match: GSS: Em:AQ406955"
complement(10624..11071)
/note="match: GSS: Em:AQ123242"
11814..11969
/note="4 copies 39 mer 90% conserved"
12082..12147
/note="LTR41 repeat: matches 90..159 of consensus"
complement(12359..13029)
/note="match: GSS: Em:AQ375128"
complement(12653..13029)
/note="match: STS: Em:AB192021"
12846..13310
/note="HERVL repeat: matches 4662..5115 of consensus"
13606..14459
/note="HERVL repeat: matches 3756..4662 of consensus"
14701..14847
/note="HERVL repeat: matches 3340..3476 of consensus"
15150..15608
/note="HERVL repeat: matches 2831..3340 of consensus"

```

repeat_region 15704..16013
/note="HERV.L repeat: matches 2343..2665 of consensus"
repeat_region 19019..19537
/note="MUT2FA repeat: matches 1..447 of consensus"
misc_feature 21461..21758
/note="match: GSS: Em:AQ094630"
misc_feature 23017..23418
/note="match: GSS: Em:AQ439345"
repeat_region 23109..23158
/note="L1MA10 repeat: matches 6215..6265 of consensus"
repeat_region 23168..23680
/note="L1MA4 repeat: matches 1..488 of consensus"
repeat_region 26002..26043
/note="21 copies 2 mer aa 83% conserved"
misc_feature complement(30902..31318)
/note="match: GSS: Em:AQ007859"
misc_feature 31310..31779
/note="match: GSS: Em:AQ229049"
misc_feature complement(31360..31705)
/note="match: GSS: Em:AQ279514"
misc_feature 31763..32205
/note="match: GSS: Em:B76328"
misc_feature complement(33002..33191)
/note="match: GSS: Em:AQ037084"
repeat_region 33228..37447
/note="L1MA5A repeat: matches 2128..6289 of consensus"
repeat_region 40468..40909
/note="L1R3 repeat: matches 5..520 of consensus"
repeat_region 42098..42151
/note="27 copies 2 mer ac 90% conserved"
misc_feature complement(46083..46507)
/note="match: GSS: Em:AQ076537"
misc_feature complement(49923..50350)
/note="match: GSS: Em:AQ115490"
repeat_region 50418..50463
/note="23 copies 2 mer tg 76% conserved"
misc_feature complement(51169..51637)
/note="match: GSS: Em:AQ221216"
repeat_region 54190..59443
/note="L1PA2 repeat: matches 889..6144 of consensus"
repeat_region 59422..60199
/note="L1PA2 repeat: matches 5..776 of consensus"
misc_feature complement(60963..61405)
/note="match: GSS: Em:AQ395505"
misc_feature complement(62363..62836)
/note="match: GSS: Em:AQ798193"
misc_feature complement(64146..64653)
/note="match: GSS: Em:AQ692409"
misc_feature complement(64181..64650)
/note="match: GSS: Em:AQ516612"
misc_feature complement(64192..64664)
/note="match: GSS: Em:AQ772035"
misc_feature complement(64204..64657)
/note="match: GSS: Em:AQ165313"
misc_feature complement(64428..64982)
/note="match: GSS: Em:AQ093975"
misc_feature 67568..67905
/note="match: GSS: Em:AQ308501"
misc_feature complement(68783..69204)
/note="match: GSS: Em:AQ513641"
repeat_region 72119..72225
/note="L1MA8 repeat: matches 6183..6288 of consensus"
repeat_region 72226..72289
/note="32 copies 2 mer aa 67% conserved"
misc_feature complement(73534..73897)
/note="match: GSS: Em:AQ050770"
repeat_region 74262..74289
/note="14 copies 2 mer ct 92% conserved"
misc_feature complement(75009..75531)
/note="match: GSS: Em:AQ36071"
misc_feature match: GSS: Em:AQ318565"
misc_feature 75556..75902
/note="match: GSS: Em:AQ142531"

```

```

misc_feature complement(81817..82461)
/note="match: GSS: Em:AQ352745"
repeat_region 81912..81993
/note="41 copies 2 mer tt 67% conserved"
misc_feature 82463..82777
/note="match: GSS: Em:B49411"
misc_feature 84199..84771
/note="match: GSS: Em:AQ085502"
misc_feature complement(89434..89956)

Query Match 72.0%; Score 21.6; DB 90; Length 101384;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 acattataggagctatgctacagata 28
Db 5174 ACATATATAAGCCATTCACAGATA 5147

RESULT 10
AL158082 LOCUS 164247 bp DNA HTG 20-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-57H12 map p21.3-22.3, ***
ACCESSION AL158082
VERSION 1.0
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 164247)
AUTHORS Pavitt,R.
JOURNAL Direct Submission
Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jul 15, 2000 this sequence version replaced gi:8217620.

COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA57H12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 159742 bases at least Q40
Consensus quality: 161714 bases at least Q30
Consensus quality: 162661 bases at least Q20
Insert size: 163347; sum-of-contigs
Insert size: 150310; 19.2% error; agarose-fp
Quality coverage: 4.32x in Q20 bases; sum-of-contigs Quality
coverage: 4.70x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 25166: contig of 25166 bp in length
* 25167 25266: gap of 100 bp
* 25267 32593: contig of 7327 bp in length
* 32594 32693: gap of 100 bp
* 32694 53020: contig of 20327 bp in length
* 53021 53120: gap of 100 bp
* 53121 94385: contig of 41265 bp in length
* 94386 94485: gap of 100 bp

```

* 94486 125966: contig of 31481 bp in length
* 125967 126066: gap of 100 bp
* 126067 130400: contig of 4334 bp in length
* 130401 130500: gap of 100 bp
* 130501 134232: contig of 3732 bp in length
* 134233 134332: gap of 100 bp
* 134333 146013: contig of 11681 bp in length
* 146014 146113: gap of 100 bp
* 146114 151697: contig of 5584 bp in length
* 151698 151797: gap of 100 bp
* 151798 164247: contig of 12450 bp in length.
Location/Qualifiers
1. 164247
/organism:"Homo sapiens"
/db_xref:"taxon:9606"
/chromosome:"1"
/map:"p21.3-22.3"
/clone:"RP11-57H12"
/clone_1b:"RPC1-11.1"
1. 25166
/note:"assembly_fragment:00994
fragment_chain:1
clone_end:T7
vector_side:left"
25267. 32593
/note:"assembly_fragment:00734
fragment_chain:1"
32694. 53020
/note:"assembly_fragment:00015
fragment_chain:1"
53121. 94385
/note:"assembly_fragment:01403
fragment_chain:1"
94486. 125966
/note:"assembly_fragment:00788
fragment_chain:1"
126067. 130400
/note:"assembly_fragment:00976"
130501. 134232
/note:"assembly_fragment:01182
fragment_chain:2"
134333. 146013
/note:"assembly_fragment:00995
fragment_chain:2"
146114. 151697
/note:"assembly_fragment:00218
fragment_chain:2"
151798. 164247
/note:"assembly_fragment:01398
fragment_chain:2
clone_end:SP6
vector_side:right"
BASE COUNT 53206 a 32822 c 29774 g 47539 t 906 others
ORIGIN
-Query Match 72.0%: Score 21.6; DB 79; Length 164247;
Best Local Similarity 85.7%: Pred. No. 80;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 acattatagagctatctgctacagata 28
||||| ||| ||||| ||||| |||||
Db 45478 ACATATATAAAGCATTCACACAGATA 45505
RESULT 11
AL356152 171713 bp DNA HTG 23-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP11-335D10, *** SEQUENCING IN
DEFINITION PROGRESS ***, 26 unordered pieces.
ACCESSION AL356152
VERSION AL356152.4 GI:9926611
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 171713)
Plumb, B.
Direct Submission
Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 25, 2000 this sequence version replaced g1:9213762.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA335D10
----- Summary Statistics
Assembly program: XGAP; Version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 158368 bases at least Q40
Consensus quality: 163406 bases at least Q30
Consensus quality: 166479 bases at least Q20
Insert size: 165213; sum-of-contigs
Insert size: 183986; 6.1% error; agarose-fp
Quality coverage: 3.30x in Q20 bases; sum-of-contigs Quality
coverage: 3.10x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 16049: contig of 16049 bp in length
* 16050 16149: gap of 100 bp
* 16150 20702: contig of 4553 bp in length
* 20703 20802: gap of 100 bp
* 20803 29599: contig of 8797 bp in length
* 29600 29699: gap of 100 bp
* 29700 33636: contig of 3837 bp in length
* 33637 33736: gap of 100 bp
* 33737 40198: contig of 6462 bp in length
* 40199 40298: gap of 100 bp
* 40299 43998: contig of 3700 bp in length
* 43999 44098: gap of 100 bp
* 44099 46584: contig of 2486 bp in length
* 46585 46684: gap of 100 bp
* 46685 48917: contig of 2233 bp in length
* 48918 49017: gap of 100 bp
* 49018 66152: contig of 17135 bp in length
* 66153 66252: gap of 100 bp
* 66253 78474: contig of 12222 bp in length
* 78475 78574: gap of 100 bp
* 78575 83439: contig of 4665 bp in length
* 83440 83539: gap of 100 bp
* 83540 92939: contig of 9400 bp in length
* 92940 93039: gap of 100 bp
* 93040 96452: contig of 3413 bp in length
* 96453 96552: gap of 100 bp
* 96553 100081: contig of 3529 bp in length
* 100082 100181: gap of 100 bp
* 100182 102600: contig of 2419 bp in length
* 102601 102700: gap of 100 bp
* 102701 105872: contig of 3172 bp in length
* 105873 105972: gap of 100 bp
* 105973 108830: contig of 2858 bp in length
* 108831 108930: gap of 100 bp
* 108931 114742: contig of 5812 bp in length


```

* 114743 .114842: gap of 100 bp
* 114843 .117118: contig of 2276 bp in length
* 117119 .117218: gap of 100 bp
* 117219 .120537: contig of 3319 bp in length
* 120538 .120637: gap of 100 bp
* 120638 .127137: contig of 6500 bp in length
* 127138 .127237: gap of 100 bp
* 127238 .131538: contig of 4301 bp in length
* 131539 .131638: gap of 100 bp
* 131639 .155611: contig of 23973 bp in length
* 155612 .155711: gap of 100 bp
* 155712 .164083: contig of 8372 bp in length
* 164084 .164183: gap of 100 bp
* 164184 .169279: contig of 5096 bp in length
* 169280 .169380: gap of 100 bp
* 169380 .171713: contig of 2334 bp in length.
Location/Qualifiers
1. .171713
Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-335D10"
/clone_lib="RP11-11.2"
1. .16049
misc_feature
/note="assembly_fragment:01329
fragment_chain:1"
16150. .20702
misc_feature
/note="assembly_fragment:00147
fragment_chain:1"
20803. .29599
misc_feature
/note="assembly_fragment:00023
fragment_chain:1"
29700. .33636
misc_feature
/note="assembly_fragment:00460
fragment_chain:1"
33737. .40198
misc_feature
/note="assembly_fragment:00174
fragment_chain:1"
40299. .43998
misc_feature
/note="assembly_fragment:00581
fragment_chain:1"
44099. .46584
misc_feature
/note="assembly_fragment:01292
fragment_chain:1"
46685. .48917
misc_feature
/note="assembly_fragment:00288
fragment_chain:2"
49018. .66152
misc_feature
/note="assembly_fragment:01057
fragment_chain:2"
66253. .78474
misc_feature
/note="assembly_fragment:00366
fragment_chain:2"
78575. .83439
misc_feature
/note="assembly_fragment:00812
fragment_chain:2"
83540. .92939
misc_feature
/note="assembly_fragment:00467
fragment_chain:2"
93040. .96452
misc_feature
/note="assembly_fragment:00273
fragment_chain:3"
96553. .100081
misc_feature
/note="assembly_fragment:00956
fragment_chain:3"
100182. .102600
misc_feature
/note="assembly_fragment:00455
fragment_chain:4"
102701. .105872
misc_feature
/note="assembly_fragment:00381
fragment_chain:4"
105973. .108830
misc_feature
/note="assembly_fragment:00672

```

```

misc_feature
fragment_chain:5"
108931. .114742
/note="assembly_fragment:00929
fragment_chain:5"
114843. .117118
misc_feature
/note="assembly_fragment:00764
fragment_chain:6"
117219. .120537
misc_feature
/note="assembly_fragment:00877
fragment_chain:6"
120638. .127137
misc_feature
/note="assembly_fragment:01242
fragment_chain:7"
127238. .131538
misc_feature
/note="assembly_fragment:01124
fragment_chain:7"
131639. .155611
misc_feature
/note="assembly_fragment:00298"
155712. .164083
misc_feature
/note="assembly_fragment:00580"
164184. .169279
misc_feature
/note="assembly_fragment:00944"
169380. .171713
misc_feature
/note="assembly_fragment:01198
vector_side:right"
BASE COUNT 54105 a 32620 c 32488 g 49975 t 2525 others
ORIGIN
Query Match 72.0%; Score 21.6; DB 80; Length 171713;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 acattataggagctatgctacagata 28
Db 92623 ACATATATATAAGCCATGCTACAGATA 92650
RESULT 12
AF283999 2785 bp DNA MAM 05-MAR-2001
LOCUS
DEFINITION Dasybus novemcinctus BRCA1 gene, partial cds.
ACCESSION AF283999
VERSION AF283999.1 GI:13195172
KEYWORDS
SOURCE
ORGANISM
Dasybus novemcinctus
nine-banded armadillo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Edentata; Dasypodidae; Dasybus.
REFERENCE
1 (bases 1 to 2785)
Madsen,O., Scally,M., Donady,C.J., Kao,D.J., Debry,R.W., Adkins,R.,
Amline,H.M., Stanhope,M.J., de Jong,W.W. and Springer,M.S.
Parallel adaptive radiations in two major clades of placental
mammals
JOURNAL
Nature 409 (6820), 610-614 (2001)
MEDLINE
21082081
REFERENCE
2 (bases 1 to 2785)
Madsen,O., Scally,M., Donady,C., Kao,D., Debry,R., Adkins,R.,
Amline,H., Stanhope,M., de Jong,W. and Springer,M.
Direct Submission
Submitted (30-JUN-2000) Biology, University of California, 1354
Spleth, Riverside, CA 92521, USA
Location/Qualifiers
1. .2785
FEATURES
source
/organism="Dasybus novemcinctus"
/db_xref="taxon:9361"
<1. .>2785
/product="BRCA1"
<1. .>2785
/note="similar to Homo sapiens BRCA1 gene exon 11"
/codon_start=2
/product="BRCA1"

```

/protein_id="AKI5586.1"
/db_xref="GI:13195173"
/translation="CGNTHANLLOHENSLLTKRDMNVEKAEFCNKSQKOPGLARQ
QSMWASKETCNDROPSTIEKYVDADPLXGKELNKKQPCSESRDQDIPWIL
NSTIOKVENFSGDILITSDSDHDSRSELAENAGALSKYKDEVSSEKIDMA
INPHDTLQFASERVQULPAESNIDKIDFTYHKKASLPKISHITRNLFICALAEK
ITQEHSLONKIKRRRTASGLRPEDLSKAYDLVYQKTPKINQTMQMDQNDVMI
ANGHENETKDCVQKEKNANPESLGEKSAFTVCGEPISSISINMGTLELINAKS
KKRRKMSSTRTHALELVGSNRPSPNHTLEIDSCSEIEEKINSNOCKPPIRNR
MLQTSKETTGAKKNNKNEPDSERHADALELKNMDELPCSSSDKLOFNSS
LQGBVANLTTIOVSDSTRPDELVSGECLOTERASESTGISVPDITDGTDSIS
LLEADITGAKATLALNOHVOYVALRNATELSHGCSKDTNDEDFKDSLREHVTG
TWEIESELDTOLOMTFKISKROSPALSPENECATVCAHSEFLQKQSPKYTEC
RHKENOGKKESKIKHVOVHTTQAGPITVCOKPDGAYAKSGIOGVRLOCSQARN
ESELINSEHEISNPQDPSLSHKMSVTKCKENLSEKEELTVSLERTVNT
IOCTVTSISHSNIRENTEKASSSINMGSDENIOAEFGRNAPKLMALRLGLMO
PEYKOSLPITNCKYPRFKSGOGEAEIRAVDIDFSCILSDINQLMGNSCAQOTS
ETPDLDLDEIKENNCFAESDIKERSAIFSKTYQKEFRSPSLVHTSFAQHQK
PKRLDSEEDVSSSE"

BASE COUNT 1049 a 499 c 571 g 666 t

ORIGIN

Query Match 71.3%; Score 21.4; DB 7; Length 2785;
Best Local Similarity 95.7%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 ttatatagactattgctacaga 26
|||||
Db 626 TTATTAGAGCTATTGCGCAGA 648

RESULT 13
LOCUS ALJ58133/c 91859 bp DNA PRI 20-NOV-2000
DEFINITION Human DNA sequence from clone RP11-810122 on chromosome 6, complete
sequence.
ACCESSION ALJ58133
VERSION ALJ58133.11 GI:11322005
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 91859)
PHLIMORE,B.
Direct Submission
Submitted (19-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Nov 23, 2000 this sequence version replaced gi:11225767.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

COMMENT

This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-810122 is from the library RPCI-11.3 constructed at the

Roswell Park Cancer Institute by the group of Pieter de Jong. For
further details see <http://bacpac.med.buffalo.edu/>
VECTOR: PBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-810122. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-810122 is at 91859 in this
sequence. The true right end of clone RP3-375M21 is at 100 in this
sequence.

FEATURES

source
1..91859
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-810122"
/clone_11p="RPCI-11.3"
/complement(1..102)
/note="match: GSS: Em:AQ014142"
744..811
/note="17 copies 4 mer atat 72% conserved"
745..812
/note="34 copies 2 mer ta 72% conserved"
1917..2342
/note="MLR1C repeat: matches 26..464 of consensus"
3286..3425
/note="MLR1D repeat: matches 363..505 of consensus"
3426..3872
/note="MSTRB repeat: matches 3..426 of consensus"
3873..4166
/note="MLR1D repeat: matches 83..363 of consensus"
complement(4436..4748)
/note="match: GSS: Em:AQ11995"
4538..4635
/note="MIR repeat: matches 109..212 of consensus"
4801..5192
/note="match: GSS: Em:AQ079663"
4833..5299
/note="match: GSS: Em:AQ307374"
5242..5373
/note="MIR repeat: matches 132..257 of consensus"
complement(5879..6511)
/note="match: GSS: Em:AQ389784"
6790..7089
/note="Alusq repeat: matches 1..302 of consensus"
7092..7226
/note="MIR repeat: matches 14..153 of consensus"
7391..7547
/note="MLR1F repeat: matches 177..338 of consensus"
7548..7863
/note="Alusp repeat: matches 1..313 of consensus"
7864..8048
/note="MLR1F repeat: matches 1..177 of consensus"
8325..8892
/note="LTR18B repeat: matches 2..592 of consensus"
8898..9624
/note="LTR39 repeat: matches 1..794 of consensus"
9950..10154
/note="MER20 repeat: matches 1..214 of consensus"
10545..10851
/note="Alusx repeat: matches 1..299 of consensus"
10873..11038
/note="MIR repeat: matches 94..252 of consensus"
13553..13800
/note="L1MA7 repeat: matches 5954..6214 of consensus"
13855..14017
/note="FAM repeat: matches 1..168 of consensus"
14253..14463
/note="L1PA8 repeat: matches 5949..6160 of consensus"
complement(14581..15134)
/note="match: GSS: Em:AQ482234"
14662..14785
/note="62 copies 2 mer gt 79% conserved"
14670..14785

misc-feature /note="29 copies 4 mer g1gt 81% conserved"
 complement(14691..15151)
 /note="match: GSS: Em:AQ393734"
 repeat_region 14915..16092
 /note="L1M4 repeat: matches 3081..4379 of consensus"
 repeat_region 16093..16158
 /note="L1M1 repeat: matches 6267..6332 of consensus"
 repeat_region 16159..17016
 /note="L1M4 repeat: matches 4379..5212 of consensus"
 repeat_region 17027..17270
 /note="L1M1 repeat: matches 109..349 of consensus"
 repeat_region 17265..17709
 /note="L1M3 repeat: matches 6219..6660 of consensus"
 repeat_region 17710..18073
 /note="L1M3 repeat: matches 1..371 of consensus"
 repeat_region 18074..19140
 /note="L1M3 repeat: matches 6660..7739 of consensus"
 repeat_region 19218..19552
 /note="L1M3 repeat: matches 5263..5576 of consensus"
 repeat_region 19574..20003
 /note="L1M3 repeat: matches 5710..6155 of consensus"
 repeat_region 20026..20056
 /note="L1M3/D repeat: matches 5434..5464 of consensus"
 repeat_region 20181..20570
 /note="L1M3 repeat: matches 1..390 of consensus"
 repeat_region 20571..20582
 /note="L1M3/D repeat: matches 5467..5601 of consensus"
 repeat_region 21348..21976
 /note="L1M3/D repeat: matches 21..460 of consensus"
 repeat_region 21996..22711
 /note="L1M3 repeat: matches 6671..7412 of consensus"
 repeat_region 22795..23848
 /note="L1M3 repeat: matches 1934..2914 of consensus"
 repeat_region 23849..24631
 /note="L1M3 repeat: matches 1..780 of consensus"
 repeat_region 24632..25014
 /note="L1M3 repeat: matches 1537..1934 of consensus"
 repeat_region 25492..25790
 /note="L1M3 repeat: matches 1..310 of consensus"
 repeat_region 26219..26762
 /note="L1M3 repeat: matches 7101..7650 of consensus"
 repeat_region 27920..27973
 /note="L1M3 repeat: matches 123..177 of consensus"
 repeat_region 28093..28174
 /note="L1M3 repeat: matches 108..189 of consensus"
 repeat_region 28555..28702
 /note="L1M3 repeat: matches 4..186 of consensus"
 repeat_region 28968..29264
 /note="L1M3 repeat: matches 1..293 of consensus"
 repeat_region 29307..29877
 /note="L1M3 repeat: matches 8..560 of consensus"
 repeat_region 30069..30570
 /note="L1M3 repeat: matches 63..616 of consensus"
 repeat_region 31298..32448
 /note="L1M3 repeat: matches 5020..6185 of consensus"
 repeat_region 32525..32806
 /note="match: GSS: Em:AQ527057"
 repeat_region 32631..32799
 /note="L1M3 repeat: matches 12..157 of consensus"
 repeat_region 33044..33355
 /note="L1M3 repeat: matches 1..309 of consensus"
 repeat_region 33562..33608
 /note="L1M3 repeat: matches 1..49 of consensus"
 repeat_region 33609..33708
 /note="L1M3 repeat: matches 6183..6286 of consensus"
 repeat_region 33913..34210
 /note="L1M3 repeat: matches 1..297 of consensus"
 repeat_region 35236..35477
 /note="L1M3 repeat: matches 1953..2164 of consensus"
 repeat_region 35500..36482
 /note="L1M3 repeat: matches 1..1083 of consensus"
 repeat_region 36464..36679
 /note="L1M3 repeat: matches 1257..1500 of consensus"

repeat_region 36680..36978
 /note="L1M3 repeat: matches 3..301 of consensus"
 repeat_region 36979..37428
 /note="L1M3 repeat: matches 1500..1887 of consensus"
 misc-feature 37989..38509
 /note="match: GSS: Em:AQ540521"
 repeat_region 38094..38248
 /note="L1M3 repeat: matches 17..199 of consensus"
 repeat_region 40216..40284
 /note="L1M3 repeat: matches 65..136 of consensus"
 repeat_region 40740..40827
 /note="L1M3 repeat: matches 1..371 of consensus"
 misc-feature 40777..40928
 /note="match: GSS: Em:AQ64188"
 repeat_region 40973..41008
 /note="match: GSS: Em:AQ64188"

Query Match 70.0%; Score 21; DB 90; Length 91859;
 Best Local Similarity 82.8%; Pred. No. 1.4e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 cattatagagactatgctacagatata 30
 Db 51505 CATTATAGAGACTATGCTACAGATATA 51477

RESULT 14
 AC021019 220216 bp DNA PRI 08-NOV-2000
 LOCUS Homo sapiens BAC clone RP11-496J3 from 6, complete sequence.
 DEFINITION AC021019
 ACCESSION AC021019.5 GI:10140837
 VERSION AC021019.5 GI:10140837
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 220216)
 Sulston, J.E. and Waterston, R.
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 99063792
 2 (bases 1 to 220216)
 Du, H., Dronov, K., Elliott, G., Gregory, S. and Martinka, S.
 The sequence of Homo sapiens BAC clone RP11-496J3
 Unpublished
 3 (bases 1 to 220216)
 Waterston, R.H.
 Direct Submission
 Submitted (12-JAN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 220216)
 Waterston, R.H.
 Direct Submission
 Submitted (15-SEP-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 220216)
 Waterston, R.
 Direct Submission
 Submitted (08-NOV-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Sep 15, 2000 this sequence version replaced g1:9838234.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0496003

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Paterno, M., Cateneze, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://Dacpac.med.buffalo.edu>)

VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-810I22. Actual start of this clone is at base position 1 of RP11-496J3; actual end is at base position 220216 of RP11-496J3.

The sequence H.NH0496J03 from base positions 144552 to 144902 is represented by sequence derived from PCR of clone DNA.

FEATURES

source Location/Qualifiers

1..220216

/organism:"Homo sapiens"

/db_xref:"taxon:9606"

/chromosome:"6"

/map:"6"

/clone:"RP11-496J3"

/clone_id:"RPCI-11"

15..319

/rpt_family:"Alu"

2423..2641

/rpt_family:"L1"

3240..3335

/rpt_family:"Alu"

3567..3762

/rpt_family:"Alu"

3947..4188

/rpt_family:"MER1-type"

6366..6580

/rpt_family:"L1"

6867..7409

/rpt_family:"L1"

7874..8520

/rpt_family:"L1"

8522..8782

/rpt_family:"L1"

8994..9279

/rpt_family:"Alu"

9880..10088

/rpt_family:"MER1-type"

10111..10208

/rpt_family:"L1"

10437..10627

/rpt_family:"MIR"

10634..10996

/rpt_family:"MALR"

11223..11519

/rpt_family:"Alu"

repeat_region

12036..12341

repeat_region

/rpt_family:"L1"

repeat_region

/rpt_family:"L1"

repeat_region

12628..12990

repeat_region

/rpt_family:"MALR"

repeat_region

13039..13095

repeat_region

/rpt_family:"L2"

repeat_region

13222..13289

repeat_region

/rpt_family:"L2"

repeat_region

14922..15347

repeat_region

/rpt_family:"MALR"

repeat_region

16291..16430

repeat_region

/rpt_family:"MALR"

repeat_region

16431..16887

repeat_region

/rpt_family:"MALR"

repeat_region

16888..17261

repeat_region

/rpt_family:"MALR"

repeat_region

17543..17640

repeat_region

/rpt_family:"MIR"

repeat_region

18247..18378

repeat_region

/rpt_family:"MIR"

repeat_region

19712..19794

repeat_region

/rpt_family:"MIR"

repeat_region

19795..20093

repeat_region

/rpt_family:"Alu"

repeat_region

20094..20231

repeat_region

/rpt_family:"MIR"

repeat_region

20396..20552

repeat_region

/rpt_family:"MALR"

repeat_region

20553..20668

repeat_region

/rpt_family:"Alu"

repeat_region

20869..21053

repeat_region

/rpt_family:"MALR"

repeat_region

21320..21897

repeat_region

/rpt_family:"ERV1"

repeat_region

21903..22629

repeat_region

/rpt_family:"MER1-type"

repeat_region

22955..23159

repeat_region

/rpt_family:"ERV1"

repeat_region

23550..23855

repeat_region

/rpt_family:"MER1-type"

repeat_region

23878..24024

repeat_region

/rpt_family:"Alu"

repeat_region

24680..24954

repeat_region

/rpt_family:"L2"

repeat_region

26558..26805

repeat_region

/rpt_family:"L1"

repeat_region

26860..27022

repeat_region

/rpt_family:"Alu"

repeat_region

27130..27255

repeat_region

/rpt_family:"L1"

repeat_region

27258..27468

repeat_region

/rpt_family:"L1"

repeat_region

28018..28593

repeat_region

/rpt_family:"L1"

repeat_region

28614..29097

repeat_region

/rpt_family:"L1"

repeat_region

29098..29163

repeat_region

/rpt_family:"L1"

repeat_region

29164..30021

repeat_region

/rpt_family:"L1"

repeat_region

30032..30275

repeat_region

/rpt_family:"L1"

repeat_region

30270..30618

repeat_region

/rpt_family:"L1"

repeat_region

30715..31078

repeat_region

/rpt_family:"MALR"

repeat_region

31132..31518

repeat_region

/rpt_family:"L1"

repeat_region

31519..31595

repeat_region

/rpt_family:"MALR"

repeat_region

31624..32145

repeat_region /rpt_family="L1"
32170 .32557
repeat_region /rpt_family="L1"
32579 .33008
repeat_region /rpt_family="L1"
33027 .33185
repeat_region /rpt_family="L1"
33186 .33575

Query Match 70.0%; Score 21; DB 88; Length 220216;
Best Local Similarity 82.8%; Pred. No. 1.4e+02;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cattatagagctatgctacagataca 30
||||||| ||| ||||| ||||| |||||
Db 64510 CATTATAGAGGAGATTGCTCAGATATA 64482

RESULT 15

AF284019 2773 bp DNA MAM 05-MAR-2001

LOCUS Dugong dugon BRCA1 gene, partial cds.

DEFINITION AF284019

ACCESSION AF284019.1 GI:13195212

VERSION

KEYWORDS

SOURCE

ORGANISM

dugong.
Dugong dugon
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Sirenia; Dugongidae; Dugong.

REFERENCE 1 (bases 1 to 2773)
Madsen, O., Scally, M., Donady, C. J., Kao, D. J., Debry, R. W., Adkins, R.,
Amrine, H. W., Stanhope, M. J., de Jong, W. W. and Springer, M. S.

TITLE Parallel adaptive radiations in two major clades of placental
mammals

JOURNAL Nature 409 (6820), 610-614 (2001)

MEDLINE 21082081

REFERENCE 2 (bases 1 to 2773)
Madsen, O., Scally, M., Donady, C., Kao, D., Debry, R., Adkins, R.,
Amrine, H., Stanhope, M., de Jong, W. W. and Springer, M.

TITLE Direct Submision
Submitted (30-JUN-2000) Biology, University of California, 1354
Spleth, Riverside, CA 92521, USA

FEATURES
Location/Qualifiers

1..2773
/organism="Dugong dugon"

db_xref="taxon:29137"

<1..>2773
/product="BRCA1"

<1..>2773
/note="Similar to Homo sapiens BRCA1 gene exon 11"

/codon_start=2
/product="BRCA1"

/protein_id="AK15606.1"

/db_xref="GI:13195213"

/translation="CGTNTTAASTLQHNSSLLITKDRMNVKAEFCCHKSKOPGLARSQ
QSRVAESKETCNDROTPTSEKRVKNNANPVEKRVKOKPCSESVRDODIPWITL
NSSIOKVNEMFFRSGDLDLHDGSESNAAVEEVEVGSSEKIDLMASDP
HSALICESERVLEPAESNIEDIKGTYRRKAILPHLSHVEDLIGAVATEPQIAO
ERPLNTKLKRRRTSGLPEDPIKVDLAAYQKPEKINQETDQVEQNGVNTINGC
HGNETKDDYVQREKKNAPTSEITKSAFRTKAPTSSISNMELELNHNSKAPKKNR
LRRSSTRIHALLLVNPNPSPPTHTLQIDSWSSSEIKKSSQKPYRHRNRLQLM
KNQETTGAKKSNKPEQISKRAHDAYPELNLSTGLITNCSSSHNYQERNLQLM
EIEENLGTIOVSNTRDPEDLVNGRGLQETISVESTSIVIPDTQVSGNSISLC
EADTLRKAKTAPNOCASOCALLENPNELIHGCPKTRNDTEDEKDLRCVENHIOETC
VEMEDXELDTQYLOSTFEKVSROSALFNSPEKATYCAHSKSLRKOSPKYTPPYGE
EEENQNKESKIKRQAVHTTAGTPEDCQKEKPSDYTKSSITKGVSRLOCSQSPRGE
SQHTTAGHGISQNDQMLSPIRASVSKKNLSEERPEHTISLRAVGNESIIVOS
TVSTVSQNDIKREPEIKRQGENGVQVDMDFSPCLISDNLEQPMGSSASQICSETP
YKOSLPVNCRKREPEIKRQGENGVQVDMDFSPCLISDNLEQPMGSSASQICSETP
DLDLDDDEIKENISFAESGIKERSAVFSKQDRREFRRNPSPLSHGLAQGHARGAREL
ESSDENISSE"

BASE COUNT 1029 a 490 c 611 g 642 t 1 others
ORIGIN

Query Match 69.3%; Score 20.8; DB 7; Length 2773;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 attatagagctatgctacaga 26
||||||| ||| ||||| ||||| |||||
Db 616 AATTATAGGAGCTGTTGCTACAGA 639

Search completed: September 28, 2001, 11:17:52
Job time: 4657 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 2001, 10:00:14 ; Search time 99.24 Seconds
(without alignments)
189.813 Million cell updates/sec

Title: US-09-551-645-2

Perfect score: 30
Sequence: 1 acattataggagctatgtctacagataca 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq.0601:*

1: /cgnl_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgnl_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgnl_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgnl_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgnl_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgnl_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgnl_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgnl_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgnl_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgnl_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgnl_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgnl_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgnl_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgnl_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgnl_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgnl_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgnl_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgnl_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgnl_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgnl_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgnl_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /cgnl_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28.4	94.7	5319	19	AAV35235
2	19.6	65.3	10996	20	AAV35235
3	19.4	64.7	426	21	AAV35235
4	19.4	64.7	2228	14	AAV35235
5	19.4	64.7	1664976	19	AAV35235
6	19.4	63.3	2162	22	AAV35235
7	18.8	62.7	1114	21	AAV35235
8	18.8	62.7	2094	17	AAV35235
9	18.8	62.7	2556	21	AAV35235
10	18.8	62.7	2556	21	AAV35235
11	18.8	62.7	3624	21	AAV35235

C	12	18.8	62.7	3771	21	AAV35235
C	14	18.8	62.7	3867	22	AAV35235
C	14	18.8	62.7	4225	22	AAV35235
C	15	18.8	62.7	20633	20	AAV35235
C	16	18.6	62.0	44576	21	AAV35235
C	17	18.6	62.0	44576	21	AAV35235
C	18	18.4	61.3	73	16	AAV35235
C	19	18.4	61.3	2710	18	AAV35235
C	20	18.4	61.3	4557	18	AAV35235
C	21	18.4	61.3	9086	20	AAV35235
C	22	18.2	60.7	52	18	AAV35235
C	23	18.2	60.7	380	18	AAV35235
C	24	18.2	60.7	400	18	AAV35235
C	25	18.2	60.7	580	21	AAV35235
C	26	18.2	60.7	951	21	AAV35235
C	27	18.2	60.7	1044	21	AAV35235
C	28	18.2	60.7	1167	22	AAV35235
C	29	18.2	60.7	1167	22	AAV35235
C	30	18.2	60.7	1785	20	AAV35235
C	31	18.2	60.7	9326	21	AAV35235
C	32	18.2	60.7	10813	18	AAV35235
C	33	18.2	60.7	16995	19	AAV35235
C	34	18.2	60.7	645	22	AAV35235
C	35	18.2	60.7	1550	20	AAV35235
C	36	18.2	60.7	2220	21	AAV35235
C	37	18.2	60.7	2632	22	AAV35235
C	38	18.2	60.7	3108	22	AAV35235
C	39	18.2	60.7	5817	21	AAV35235
C	40	18.2	60.7	5817	21	AAV35235
C	41	18.2	60.7	6309	21	AAV35235
C	42	18.2	60.7	1038602	20	AAV35235
C	43	17.8	59.3	585	20	AAV35235
C	44	17.8	59.3	1260	21	AAV35235
C	45	17.8	59.3	1401	14	AAV35235

ALIGNMENTS

RESULT	1
ID	AAV35235
AAV35235	standard; DNA; 5319 BP.
AC	AAV35235;
XX	
DT	28-SEP-1998 (first entry)
XX	
DE	R. prowazekii S-layer protein genomic DNA.
XX	
KW	Surface layer protein; S-layer; vaccine; typhus; spotted fever;
KW	Infection; diagnosis; disease; ds.
XX	
OS	Rickettsia prowazekii.
XX	
FH	Key
FT	-35_signal
FT	Location/Qualifiers
FT	340..345
FT	/*tag= a
FT	363..368
FT	/*tag= b
FT	379..386
FT	/*tag= c
FT	391..5229
FT	/*tag= d
FT	/*tag= e
FT	/*tag= f
FT	/*tag= g
FT	/*tag= h
FT	/*tag= i
FT	/*tag= j
FT	/*tag= k
FT	/*tag= l
FT	/*tag= m
FT	/*tag= n
FT	/*tag= o
FT	/*tag= p
FT	/*tag= q
FT	/*tag= r
FT	/*tag= s
FT	/*tag= t
FT	/*tag= u
FT	/*tag= v
FT	/*tag= w
FT	/*tag= x
FT	/*tag= y
FT	/*tag= z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag= 4
FT	/*tag= 5
FT	/*tag= 6
FT	/*tag= 7
FT	/*tag= 8
FT	/*tag= 9
FT	/*tag= A
FT	/*tag= B
FT	/*tag= C
FT	/*tag= D
FT	/*tag= E
FT	/*tag= F
FT	/*tag= G
FT	/*tag= H
FT	/*tag= I
FT	/*tag= J
FT	/*tag= K
FT	/*tag= L
FT	/*tag= M
FT	/*tag= N
FT	/*tag= O
FT	/*tag= P
FT	/*tag= Q
FT	/*tag= R
FT	/*tag= S
FT	/*tag= T
FT	/*tag= U
FT	/*tag= V
FT	/*tag= W
FT	/*tag= X
FT	/*tag= Y
FT	/*tag= Z
FT	/*tag= 0
FT	/*tag= 1
FT	/*tag= 2
FT	/*tag= 3
FT	/*tag=

PR 20-DEC-1993; 9305-0169927.
PR 09-AUG-1991; 9105-0742128.
XX
PA (USNA) US SEC OF NAVY.
XX
PI Carl M, Ching W, Dasch GA, Dobson ME;
XX
XX WPI: 1998-427031/36.
DR P-PSDB; AAM65088.
XX
PT Recombinant DNA encoding Rickettsia surface layer proteins - useful
PT for diagnosing typhus and spotted fever and for preparing vaccines
PT against them
XX
XX
PS Claim 1; Column 11-24; 20pp; English.
XX
CC This sequence encodes the Surface layer (S-layer) protein from
CC R. prowasekii strain Breinl. This sequence is useful for vaccination
CC against typhus and spotted fever rickettsial infection or for diagnosing
CC diseases caused by these bacteria. The surface layer protein antigens can
CC be produced recombinantly in large quantities.
XX
SQ Sequence 5319 BP; 1815 A; 766 C; 983 G; 1755 T; 0 other;

Query Match 94.7%; Score 28.4; DB 19; Length 5319;
Best Local Similarity 96.7%; Pred. No. 0.0054;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 acattatagaagactatgctacagataca 30
||||| ||||||| ||||||| |||||||
Db 4183 acattgtagaagactatgctacagataca 4212

RESULT 2

AAK13139/C
ID AAK13139 standard; DNA; 10996 BP.

XX
AC AAK13139;

XX
DT 19-MAR-1999 (first entry)

XX
DE Enterococcus faecalis genome contig SEQ ID NO:202.

XX
KM Enterococcus faecalis; contig; detection; Enterococcal infection;

XX
KW vaccine; attenuation; computer readable medium; ds.

XX
OS Enterococcus faecalis.

XX
PN M09850555-A2.

XX
PD 12-NOV-1998.

XX
XX 04-MAY-1998; 98WO-US08985.

XX
PR 14-NOV-1997; 97US-0066009.

XX
PR 06-MAY-1997; 97US-0044031.

XX
PR 16-MAY-1997; 97US-0046655.

XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI Barash SC, Dillon PJ, Kunsch CA;

XX
DR WPI: 1999-045171/04.

XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.

XX
PS Claim 1; Page 1065-1071; 2084pp; English.

XX
CC A computer readable medium has been developed which has recorded on it

CC 992 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAK12938 to AAK13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 10996 BP; 3633 A; 1848 C; 2288 G; 3223 T; 4 other;

Query Match 65.3%; Score 19.6; DB 20; Length 10996;
Best Local Similarity 84.6%; Pred. No. 31;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 attataagagactatgctacagata 28
||||| ||||| ||||| |||||
Db 9681 ATGTATAGTACTTTTGTCTAGAGATA 9656

RESULT 3

AAC00105/C
ID AAC00105 standard; cDNA; 426 BP.

XX
AC AAC00105;

XX
DT 06-OCT-2000 (first entry)

XX
DE Human secreted protein 5' EST, SEQ ID NO: 103.

XX
KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX
KW gene therapy; chromosome mapping; ss.

XX
OS Homo sapiens.

XX
PN EP1033401-A2.

XX
PD 06-SEP-2000.

XX
PF 21-FEB-2000; 2000EP-0200610.

XX
PR 26-FEB-1999; 99US-0122487.

XX
PA (GEST) GENSET.

XX
PI Dumas Mline Edwards J, Duclert A, Giordano J;

XX
DR WPI: 2000-500381/45.

XX
DR P-PSDB; AAG00099.

XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX
PS Claim 1; SEQ ID 103; 71pp + CD-ROM; English.

XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
XX
SQ Sequence 426 BP; 110 A; 92 C; 124 G; 100 T; 0 other;

Query Match 64.7%; Score 19.4; DB 21; Length 426;
Best Local Similarity 79.3%; Pred. No. 25;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 cattatagagactgtacagataca 30
|||||
DB 396 CATTCTAGCAGCTAATGCTGCGCTGCA 368

RESULT 4
AAQ34554/C
ID AAQ34554 standard; cDNA; 2228 BP.

XX AAQ34554;

XX 27-MAY-1993 (first entry)

XX cDNA encoding rp40-T1A-1.

XX Antigen; cytolytic lymphocyte; infection; HIV; targeting; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
FH 57..1184
FT CDS /*tag= a

XX WO9301314-A.

XX 21-JAN-1993.

XX 12-JUN-1992; 92WO-US05117.

XX 10-JUL-1991; 91US-0726607.

XX 19-FEB-1992; 92US-0843949.

XX (DAND) DANA FARBER CANCER INST INC.

XX Anderson PJ, Schlossman SF, Streuli M;

XX MPI: 1993-045515/05.

XX P-PSDB; AAR32009.

XX DNA encoding T1A-1 antigen - used for detecting and cytotoxic
XX killing of cytolytic lymphocyte(s) applicable for cancer
XX treatment

XX Claim 3; Page 35; 61pp; English.

XX RNA from cytolytic T cell clone T478C1, which expresses high levels
XX of T1A-1 antigen was used for the construction of a cDNA library in
XX lambda g11. When this expression library was screened using T1A-1
XX monoclonal antibody, several bacteriophage expressing immunoreactive
XX fusion proteins were identified. Cross hybridisation analysis
XX identified three independent clones conrg. related cDNAs. The
XX largest cDNA insert was used to probe the original library for
XX isolation of full length cDNAs. When this same probe was used to
XX screen a second lambda g11 library prepd. from PHA-activated T
XX cells, a 2.2 kb insert cDNA was isolated. clone 2G9.4, which encodes
XX rp40-T1A-1, a 40 kD T1A-1 antigen. A segment of the cDNA is useful
XX as a probe for determining the presence of cytolytic lymphocytes in a
XX biological sample. This ability provides an early warning of the
XX presence of an infective agent, e.g. HIV virus.
XX See also AAQ34555-6.

XX Sequence 2228 BP; 705 A; 385 C; 458 G; 680 T; 0 other;

Query Match 64.7%; Score 19.4; DB 14; Length 2228;
Best Local Similarity 79.3%; Pred. No. 31;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 cattatagagactgtacagataca 30
|||||
DB 258 CATTCTAGCAGCTAATGCTGCGCTGCA 230

RESULT 5
AAV21209/C
ID AAV21209 standard; DNA; 1664976 BP.

XX AAV21209;

XX 10-NOV-1998 (first entry)

XX Methanococcus jannaschli circular chromosome.

XX Methanococcus jannaschli; methanogenic archaeon; circular chromosome;
XX genome; autotrophic; extrachromosomal element; identification; ds.

XX Methanococcus jannaschli.

XX WO9607830-A2.

XX 26-FEB-1998.

XX 22-AUG-1997; 97WO-US14900.

XX 22-AUG-1996; 96US-0024428.

XX (GENO-) INST GENOMIC RES.

XX (UNIT) UNIV ILLINOIS FOUND.

XX (UJJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX Bult CJ, Smith HO, Venter JC, White OR, Woese CR;

XX MPI: 1998-169145/15.

XX Complete genome sequence of methano-genic archaeon, Methanococcus
XX jannaschli - useful in identification of M. jannaschli genome
XX fragment

XX Claim 13; Page 152-585; 614pp; English.

XX The present sequence represents the complete 1.66-megabase pair genome
XX sequence of the Methanococcus jannaschli circular chromosome. The
XX present invention describes M. jannaschli open reading frames from the
XX genome sequence. The invention also describes a computer based system
XX for identifying fragments of the M. jannaschli genome that are
XX homologous to target nucleotide sequences, comprising: (a) data storage
XX means comprising the nucleotide sequence of the 1664976, 58407 or 16550
XX bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide
XX sequence at least 99.9% identical to it; (b) search means for comparing a
XX target sequence to the nucleotide sequence of the data storage means to
XX identify a homologous sequence, and (c) retrieval means for obtaining
XX the homologous sequence. The method, which is based on whole genome
XX random sequencing of an autotrophic archaeon M. jannaschli, the genome
XX of which consists of 3 physically distinct elements, a large circular
XX chromosome (the 1664976 bp sequence given in AAV21209), a large circular
XX extra-chromosomal element (the 58407 bp sequence given in AAV21210), and
XX a small circular extra-chromosomal element (the 16550 bp sequence given
XX in AAV21211), can be used in the identification of M. jannaschli genome
XX fragment.

XX Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;

Query Match 64.7%; Score 19.4; DB 19; Length 1664976;
Best Local Similarity 79.3%; Pred. No. 66;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 acattatagagctattgctacagatca 29
 | |||| | |||| |||| |||| ||
 Db 388753 AGATTTCAGTACGCTATTGCTACAGTAAC 388725

RESULT 6
 AAC60223
 ID AAC60223 standard; cDNA; 2162 BP.
 AC AAC60223;
 DT 13-FEB-2001 (first entry)
 XX
 XX Human hydrolase-like molecule 1 cDNA.
 XX
 XX Hydrolase-like molecule; human; cell proliferation disorder;
 KW autoimmune; cancer; AIDS; acquired immune deficiency syndrome; ss.
 XX
 XX Homo sapiens.
 OS
 PN US6132964-A.
 PD 17-OCT-2000.
 PF 06-FEB-1998; 98US-0013881.
 PR 06-FEB-1998; 98US-0013881.
 PA (INCY-) INCYTE PHARM INC.
 PI Bandman O, Hillman JL, Guegler KJ, Shah P, Lai P, Corley NC;
 DR WPI: 2001-006133/01.
 DR P-PSDB; AAB28790.
 XX
 PT New human hydrolase-like molecules (HHLMs) and polynucleotides encoding
 PT the HHLMs, useful for diagnosing, treating or preventing cell
 PT proliferation (e.g. bone cancer) or autoimmune disorders(e.g. AIDS or
 PT asthma) -
 XX
 PS Claim 5; Column 57-58; 38pp; English.
 CC The present invention relates to isolated and purified cDNA encoding a
 CC human hydrolase-like molecule (HHLM), designated HHLM-1 to HHLM-8. The
 CC HHLM DNAs and polypeptides are useful for diagnosing, treating or
 CC preventing cell proliferation disorders and autoimmune disorders. Cell
 CC proliferation disorders include cancers, autoimmune disorders include
 CC AIDS (acquired immune deficiency syndrome). The present sequence
 CC is a cDNA encoding a HHLM protein of the invention.
 XX
 SQ Sequence 2162 BP; 629 A; 383 C; 508 G; 642 T; 0 other;

Query Match 63.3%; Score 19; DB 22; Length 2162;
 -Best Local Similarity 81.5%; Pred. No. 45;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 4 ttatagagctattgctacagatca 30
 ||| ||| |||| |||| |||| ||
 Db 1815 ttatagagctattgctacagatca 1841

RESULT 7
 AA229770/c
 ID AA229770 standard; DNA; 1114 BP.
 AC AA229770;
 XX
 XX 27-MAR-2000 (first entry)
 DT
 KW Potato alpha-amylase gene promoter.
 DE

XX
 KW Potato alpha-amylase promoter sequence; alpha-amylase;
 KW phenotype modulating genetic sequence; PMGS; transposon tagging;
 KW ds element; dissociation element; UQ406 sequence; starch metabolism;
 KW plant pathogen resistance; senescence timing; cell growth; ds.
 XX
 OS Solanum tuberosum.
 XX
 XX Key Location/Qualifiers
 FH 647..654
 FT transposon
 FT /*tag= a
 FT /note= "UQ406 insertion with single Ds element"
 XX
 XX W09963068-A1.
 XX
 XX 09-DEC-1999.
 XX
 XX 04-JUN-1999; 99WO-AU00434.
 XX
 XX 04-JUN-1998; 98AU-0003901.
 PR 04-JUN-1998; 98AU-0003903.
 PR 25-SEP-1998; 98AU-0006169.
 PR 25-SEP-1998; 98AU-0006174.
 XX
 PA (UYQU) UNIV QUEENSLAND.
 XX
 XX Carroll BJ;
 PI
 DR WPI: 2000-116368/10.
 XX
 XX
 PT New polynucleotides that increase gene expression in plants used to
 PT produce transgenic plants with resistance to plant pathogens -
 XX
 PS Claim 20; Fig 4; 93pp; English.
 XX
 CC The present sequence is potato alpha-amylase promoter sequence. This
 CC shows strong similarity to an alpha-amylase promoter of tomato. The
 CC manipulation of amylase levels is done by introducing isolated phenotype
 CC modulating genetic sequence which increases or stabilizes expression of a
 CC second nucleotide sequence inserted proximally. Genomewalker (14) is used
 CC to clone tomato DNA sequences flanking the Ds element in UQ406. The Ds
 CC sequence inserts into the promoter region. This is used in transposon
 CC tagging of alpha-amylase gene to identify mutants exhibiting altered
 CC physiological properties. Transgenic plants having altered phenotypic
 CC traits, such as resistance to plant pathogens, senescence timing, starch
 CC metabolism, cell growth, expansion and/or division, and the shape of
 CC cells, tissues or organs can be produced.
 XX
 SQ Sequence 1114 BP; 326 A; 207 C; 178 G; 403 T; 0 other;

Query Match 62.7%; Score 18.8; DB 21; Length 1114;
 Best Local Similarity 76.7%; Pred. No. 51;
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 acattatagagctattgctacagatca 30
 || |||| || |||| |||| ||||
 Db 45 ACTCTATATAGTTTCTAATGCTACAGATATA 16

RESULT 8
 AAT30126/c
 ID AAT30126 standard; DNA; 2094 BP.
 AC AAT30126;
 XX
 XX 08-JAN-1997 (first entry)
 DT
 DE Alpha-amylase 1 promoter.
 XX
 KW Alpha-amylase 1; amy 1; promoter; plant; enzyme; starch breakdown; tuber;
 KW reducing sugar; potato; sprout; stem tissue; dicotyledonous plant;
 KW agrobacterium; crop protein; nutrition; mammal; interferon; insulin;

KW blood factor; plasminogen activator; ss.
 XX Solanum tuberosum.
 OS Solanum tuberosum.
 XX Key Location/Qualifiers
 FH CAAT_signal 1783..1786 /tag= a
 FT CAAT_signal 1863..1866 /tag= b
 FT CAAT_signal 1890..1895 /tag= c
 FT TATA_signal 1906..1911 /tag= d
 FT TATA_signal 1927..1932 /tag= e
 FT TATA_signal 1927..1932 /tag= e
 PN W09612813-A1.
 XX 02-MAY-1996.
 PD 06-JUN-1995; 95WO-EP02195.
 XX 21-OCT-1994; 94GB-0021292.
 PR (DANI-) DANISCO AS.
 XX Kreiberg JD, Lund M, Pedersen HF, Pedersen RF;
 PI MPI; 1996-230612/23.
 DR WPI; 1996-230612/23.
 XX Plant alpha-amylase gene promoter - capable of expressing a gene of
 PT interest in sprout or stem tissue of a dicotyledonous plant
 XX Claim 1; Page 33-36; 48pp; English.
 PS This sequence represents the Solanum tuberosum alpha-amylase 1 promoter.
 CC This alpha-amylase is one of the key plant enzymes, and participates in the
 CC pathway responsible for the breakdown of starch to reducing sugars in
 CC potato tubers. Alpha-amylase is encoded by a gene family consisting of
 CC at least five individual genes divided into two sub families (the type 1
 CC and type 3 alpha-amylases). Type 1 alpha-amylases are expressed in
 CC sprout and stem tissues, but not in tubers, roots or leaves. The type 3
 CC alpha-amylases are expressed in tubers, sprouts and stem tissues. The
 CC promoter is used to direct expression of a gene of interest in stem
 CC cells, tissues or organs of a dicotyledonous plant (such as a potato).
 CC This sequence can be used in constructs, such as vectors (such as those
 CC derived from agrobacterium), for transforming cells to express a gene of
 CC interest. The gene of interest that is fused to this promoter sequence
 CC may be a crop protein gene with an optimised amino acid composition, so
 CC as to increase the nutritive value of the crop. It is also possible to
 CC use this promoter to express non-plant genes for mammalian products, such
 CC as interferons, insulin, blood factors, and plasminogen activators.
 CC Sequence 2094 BP; 618 A; 357 C; 369 G; 750 T; 0 other;
 .SQ
 QY 1 acattataggagctattgctacagatata 30
 DB 1025 ACTCTTATAGTTTCTATACCTACAGATATA 996
 RESULT 9
 AA02872/c
 ID AA02872 standard; cDNA; 2556 BP.
 AC AA02872;
 XX 28-APR-1998 (first entry)
 DT
 XX

DE Human hTAFII105 cDNA.
 XX TATA-binding protein associated binding factor 105; human; activator;
 KW hTAFII105; transcription factor; TFIID; transcriptional activation;
 KW antibodies; diagnosis; therapy; biopharmaceutical industry; ds.
 XX Homo sapiens.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 1..2406 /tag= a
 FT /product= hTAFII105
 FT /note= "human tata binding protein associated
 FT factor 105 partial sequence"
 PN US5710025-A.
 XX 20-JAN-1998.
 PD 02-OCT-1996; 96US-0725012.
 XX 02-OCT-1996; 96US-0725012.
 PR (REGC) UNIV CALIFORNIA.
 XX Dikstein R, Tjian R;
 PI MPI; 1998-109818/10.
 DR P-PSDB; AAW31494.
 XX DNA encoding human tata-binding protein associated factor - for
 PT producing recombinant protein
 XX Claim 7; Col 11-18; 12pp; English.
 PS This cDNA sequence encodes a human tata-binding protein associated
 CC factor, hTAFII105. Isolated from Daudi cell nuclear extracts. Tightly
 CC associated subunits (TAF's) are components of the transcription factor
 CC TFIID and are thought to mediate transcriptional activation. This encoded
 CC protein may be produced recombinantly from transformed host cells or
 CC purified from human cells. hTAFII105 specific binding agents such as
 CC specific antibodies could be used for diagnosis (e.g. genetic
 CC hybridisation screens for hTAFII105 transcripts), therapy (e.g. gene
 CC therapy to modulate hTAFII105 gene expression) and in the
 CC biopharmaceutical industry (e.g. as immunogens, reagents for isolating B
 CC cell specific activators or other transcriptional regulators).
 CC Sequence 2556 BP; 797 A; 614 C; 514 G; 631 T; 0 other;
 .SQ
 QY 1 acattataggagctattgctacagatata 30
 DB 974 ACAGTTCGGGTCTGTCTGCTCCAGAAACA 945
 RESULT 10
 AA290465/c
 ID AA290465 standard; DNA; 2556 BP.
 AC AA290465;
 XX 06-JUN-2000 (first entry)
 DT Transcription factor subunit TAFII105 polypeptide encoding DNA.
 XX TATA box-binding protein associated factor II 105; TAFII105; cancer;
 KW transcription factor; apoptosis; cytostatic; immunosuppressive;
 KW antiinflammatory; virucide; antibacterial; ds.
 XX

```
OS Homo sapiens.
XX Key Location/Qualifiers
FH CDS 1..2556
FT /*tag= a
FT /transl_except= (pos: 2404..2406, aa: GLX)
FT /transl_except= (pos: 2499..2501, aa: GLX)
FT /note= "the stop codon is not indicated"
XX
XX WO200012699-A1.
XX
XX PD 09-MAR-2000.
XX
XX PF 25-AUG-1999; 99WO-IL00464.
XX
XX PR 27-AUG-1998; 98IL-0125971.
XX
XX PA (YEDA ) YEDA RES & DEV CO LTD.
XX
XX PI Dikstein R, Yamlt-hezl A;
XX
XX DR WPI: 2000-256640/22.
XX
XX DR P-PSDB: AAY57279.
XX
XX PF Polypeptide encoding TATPA box binding protein associated factor II 105
XX useful for treating e.g. cancers and inducing apoptosis has a dominant
XX negative effect on the normal biological activity of the binding
XX protein -
XX
XX PS Claim 12; Fig 1; 48bp; English.
XX
XX CC This DNA encodes a polypeptide comprising a (modified) fragment (I) of
XX a TATA box-binding protein associated factor II 105 (TAFII105). A
XX pharmaceutical composition comprising (I) or the polynucleotide or an
XX inhibitor or antagonist of (I) is useful for treating cancers and
XX inducing apoptosis in pathological cells. The composition is also useful
XX for treating autoimmune diseases, inflammatory processes and viral or
XX bacterial infections.
XX
XX SQ Sequence 2556 BP; 797 A; 614 C; 514 G; 631 T; 0 other;
XX
XX Query Match 62.7%; Score 18.8; DB 21; Length 2556;
XX Best Local Similarity 76.7%; Pred. No. 56;
XX Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
XX
XX Oy 1 acattatagagcctatgctacagataca 30
XX ||||| ||||| ||||| ||||| |||||
XX Db 974 ACAGTTCTGGGTGCTGTCACAGAAC 945
XX
XX RESULT 11
XX AA251928/c
XX ID AA251928 standard; DNA: 3624 BP.
XX
XX AC AA251928;
XX
XX XX 04-JUL-2000 (first entry)
XX
XX DE Human latent transforming growth factor-beta binding protein 3 (I) gene.
XX
XX KW Human latent transforming growth factor-beta binding protein 3; hLTBP-3;
XX TGF-beta inhibitor; proliferative; anti-proliferative; cytostatic;
XX cardiant; anti-inflammatory; cerebroprotective; immunosuppressive;
XX thrombolytic; osteopathic; vulnery; tranquilizer; antibacterial; PAI-1;
XX plasminogen activator inhibitor-1; thrombomodulin; tissue growth; tumor;
XX prevention; treatment; cancer; fibrosis; dilated cardiomyopathy; injury;
XX osteoporosis; myocardial infarction; congestive heart failure; sepsis;
XX thrombosis; stroke; systemic inflammatory response syndrome;
XX septic shock; sepsis syndrome; multiple organ dysfunction syndrome;
XX atherosclerotic plaque rupture; ds.
XX
XX OS Homo sapiens.
```

```
XX Key Location/Qualifiers
FH CDS 1..3624
FT /*tag= a
FT /product= "hLTBP-3"
FT /note= "Does not include stop codon"
FT /partial
XX
XX WO200012551-A1.
XX
XX PD 09-MAR-2000.
XX
XX PF 30-AUG-1999; 99WO-US19436.
XX
XX PR 01-SEP-1998; 98US-0098766.
XX
XX PA (ELIL ) LILLY & CO ELI.
XX
XX PI Edmonds BT;
XX
XX DR WPI: 2000-256589/22.
XX
XX DR P-PSDB: AAY70551.
XX
XX PF Human latent transforming growth factor (TGF)-beta binding protein 3,
XX nucleic acids and vectors useful as modulators of TGF-beta, for
XX inhibiting tissue or tumor growth, and treating e.g. osteoporosis and
XX myocardial infarction -
XX
XX PS Claim 3; Page 60-61; 78bp; English.
XX
XX CC The present sequence is a gene encoding human latent transforming growth
XX factor (TGF)-beta binding protein 3 (hLTBP-3). The gene was obtained by
XX reverse transcriptase PCR of mRNA isolated from hLTBP-3 expressing
XX tissues e.g. cartilage, bone, brain, kidney, ovary, pancreas, stomach
XX and spleen. The protein is often co-expressed with TGF-beta and functions
XX as its inhibitor. hLTBP-3 is useful for inhibiting or stimulating tissue
XX growth in vitro or in vivo and for inhibiting tumor growth. The hLTBP-3,
XX or its homologues, and antisense nucleic acid sequences can be used to
XX regulate TGF-beta activity, especially plasminogen activator inhibitor-1
XX expression, activity or secretion, thrombomodulin expression or
XX activity, TGF-beta secretion and cellular proliferation. Modulation of
XX hLTBP-3 is useful for prevention and/or treatment of diseases arising
XX from cellular effects induced by TGF-beta, especially cancer, fibrosis,
XX osteoporosis, myocardial infarction, congestive heart failure, dilated
XX cardiomyopathy, deep venous thrombosis, disseminated intravascular
XX coagulopathy, stroke, sepsis, injuries involving major tissue damage and
XX trauma, systemic inflammatory response syndrome, sepsis syndrome,
XX septic shock, multiple organ dysfunction syndrome, and atherosclerotic
XX plaque rupture.
XX
XX SQ Sequence 3624 BP; 652 A; 1248 C; 1149 G; 575 T; 0 other;
XX
XX Query Match 62.7%; Score 18.8; DB 21; Length 3624;
XX Best Local Similarity 76.7%; Pred. No. 59;
XX Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
XX
XX Oy 1 acattatagagcctatgctacagataca 30
XX ||||| ||||| ||||| ||||| |||||
XX Db 2406 ACATCTGAGAGCCATTGTTATGATCA 2377
XX
XX RESULT 12
XX AA251929/c
XX ID AA251929 standard; DNA: 3771 BP.
XX
XX AC AA251929;
XX
XX XX 04-JUL-2000 (first entry)
XX
XX DE Human latent transforming growth factor-beta binding protein 3 (II) gene.
XX
XX KW Human latent transforming growth factor-beta binding protein 3; hLTBP-3;
XX
```


PN WO20010672-A1.
XX
PD 04-JAN-2001.
XX
PF 29-JUN-2000; 2000WO-US18184.
XX
PR 29-JUN-1999; 99US-0342687.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;
DR WPI; 2001-050127/06.
XX
XX Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
PT MANGO 245 proteins, useful in the treatment of inflammatory diseases
PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
PT disorders (e.g. jaundice) -
XX
PS Claim 1: Fig 9: 262pp; English.
XX
CC The present invention relates to cDNAs encoding TANGO 244,
CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.
CC The nucleic acids, proteins and protein modulators are useful for
CC treating colonic disorders, inflammatory diseases, tumors,
CC renal disorders, liver disorders, lung disorders, autoimmune diseases,
CC allergic diseases, cardiovascular diseases, brain disorders,
CC degenerative diseases placental, pancreatic, skeletal and muscle
CC disorders.
XX
SQ Sequence 4225 BP; 731 A; 1470 C; 1359 G; 665 T; 0 other;
XX
Query Match 62.7%; Score 18.8; DB 22; Length 4225;
Best Local Similarity 76.7%; Pred. No. 60;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 acattatagagctatgctacagataca 30
DB 2572 ACATCTGTAGAGCCATTGCTATGATGCA 2543
XX
RESULT 15
AA13213
ID AA13213 standard; DNA: 20633 BP.
XX
AC AA13213;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:276.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal Infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PM WO9850555-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 1999-045171/04.
XX
XX

PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1: Page 1284-1294; 2084pp; English.
PS
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AA12938 to AA13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 20633 BP; 5857 A; 4284 C; 3025 G; 7427 T; 40 other;
XX
Query Match 62.7%; Score 18.8; DB 20; Length 20633;
Best Local Similarity 76.7%; Pred. No. 74;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 acattatagagctatgctacagataca 30
DB 14803 acattatagacatctaattgcccagataca 14832
XX
Search completed: September 28, 2001, 10:26:07
JOB Time: 1553 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 2001, 10:00:14 ; Search time 45.3 Seconds
(without alignments)
125.372 Million cell updates/sec

Title: US-09-551-645-2

Sequence: 1 acattatagagctatgtctacagatata 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn1_7/ptodata/1/lna/5A.COMB.seq:*
2: /cgn1_7/ptodata/1/lna/5B.COMB.seq:*
3: /cgn1_7/ptodata/1/lna/6A.COMB.seq:*
4: /cgn1_7/ptodata/1/lna/6B.COMB.seq:*
5: /cgn1_7/ptodata/1/lna/PCRTUS.COMB.seq:*
6: /cgn1_7/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28.4	94.7	5319	1	US-08-169-927-1 Sequence 1, Appl1
2	19.4	64.7	2228	1	US-07-726-607C-1 Sequence 1, Appl1
3	19.4	64.7	2228	1	US-07-843-949A-1 Sequence 1, Appl1
4	19.4	64.7	2228	2	US-08-218-978-1 Sequence 1, Appl1
5	19.4	64.7	2228	2	US-09-013-881-9 Sequence 9, Appl1
6	18.8	62.7	2556	3	US-08-725-012-1 Sequence 1, Appl1
7	17.8	59.3	1401	2	US-07-843-949A-3 Sequence 3, Appl1
8	17.8	59.3	1401	2	US-08-218-978-3 Sequence 3, Appl1
9	17.4	58.0	1861	1	US-08-246-403A-13 Sequence 13, Appl1
10	17.4	58.0	2119	3	US-08-714-918-42 Sequence 42, Appl1
11	17.4	58.0	2119	4	US-09-265-315-42 Sequence 42, Appl1
12	17.4	58.0	2119	4	US-09-265-315-42 Sequence 42, Appl1
13	17.4	58.0	2119	4	US-09-265-315-42 Sequence 42, Appl1
14	17.4	58.0	2119	4	US-09-265-315-42 Sequence 42, Appl1
15	17.4	58.0	2119	4	US-09-265-315-42 Sequence 42, Appl1
16	17.4	58.0	2119	4	US-09-265-315-42 Sequence 42, Appl1
17	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
18	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
19	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
20	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
21	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
22	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
23	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
24	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
25	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
26	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1
27	17.2	57.3	1020	1	US-08-449-315-7 Sequence 7, Appl1

ALIGNMENTS

28	17.2	57.3	1020	2	US-08-971-217-7	Sequence 7, Appl1
29	17.2	57.3	3753	5	PCT-US95-02251-2	Sequence 2, Appl1
30	17.2	57.3	3753	3	US-08-479-722B-3	Sequence 3, Appl1
31	17.2	57.3	4314	1	US-08-199-780-2	Sequence 2, Appl1
32	17.2	57.3	4314	2	US-08-316-650-2	Sequence 2, Appl1
33	16.8	56.0	423	1	US-08-470-179-77	Sequence 77, Appl1
34	16.8	56.0	720	3	US-09-479-309-3	Sequence 3, Appl1
35	16.8	56.0	4249	1	US-08-480-784-21	Sequence 21, Appl1
36	16.8	56.0	4249	1	US-08-483-553-21	Sequence 21, Appl1
37	16.8	56.0	4249	1	US-08-487-002-21	Sequence 21, Appl1
38	16.8	56.0	4249	1	US-08-483-554B-21	Sequence 21, Appl1
39	16.8	56.0	4249	1	US-08-488-011B-21	Sequence 21, Appl1
40	16.8	56.0	4249	4	US-08-850-727-21	Sequence 21, Appl1
41	16.8	56.0	4249	5	PCT-US95-10202-21	Sequence 21, Appl1
42	16.8	56.0	4249	5	PCT-US95-10203-21	Sequence 21, Appl1
43	16.8	56.0	4249	5	PCT-US95-10220-21	Sequence 21, Appl1
44	16.8	56.0	5656	1	US-08-425-061-1	Sequence 1, Appl1
45	16.8	56.0	5656	1	US-08-825-886-1	Sequence 1, Appl1

RESULT 1
US-08-169-927-1
Sequence 1, Application US/08169927
Patent No. 5783441
GENERAL INFORMATION:
APPLICANT: Carlson, Mitchell
APPLICANT: Dobson, Michael E.
APPLICANT: Dasch, Gregory A.
TITLE OF INVENTION: Gene and Protein Applicable to the
TITLE OF INVENTION: Preparation of Vaccines for Rickettsia typhi and the Detection of Both
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical R & D Command
STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,927
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5319 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:

ORGANISM: Rickettsia prowazekii
STRAIN: Breinl
FEATURE:
NAME/KEY: -35_signal
LOCATION: 340..345
FEATURE:
NAME/KEY: -10_signal
LOCATION: 363..368
FEATURE:
NAME/KEY: CDS
LOCATION: 391..5226
FEATURE:
NAME/KEY: RBS
LOCATION: 379..386
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5270..5306
PUBLICATION INFORMATION:
AUTHORS: Carl, M.
AUTHORS: Dobson, M. E.
AUTHORS: Ching, W. M.
AUTHORS: Dasch, G. A.
TITLE: Characterization of the gene encoding the
TITLE: protective S-layer protein of Rickettsia
TITLE: prowazekii; presence of a truncated identical
TITLE: homolog in rickettsia typhi
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
DATE: 1990
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
US-08-169-927-1

Query Match 94.7%; Score 28.4; DB 1; Length 5319;
Best Local Similarity 96.7%; Pred. No. 0.002;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 acattatagagctattgctacagataca 30
||||| ||||||| ||||||| |||||||
Db 4183 ACATTGTAGAGCTATTGCTACAGATACA 4212

RESULT 2
US-07-726-607C-1/c
Sequence 1, Application US/07726607C
Patent No. 5298407
GENERAL INFORMATION:
APPLICANT: Anderson, Paul J.
APPLICANT: Streuli, Michel
APPLICANT: Schlossman, Stuart F.
TITLE OF INVENTION: DNA ENCODING A PROTEIN
TITLE OF INVENTION: ACTIVE IN LYMPHOCYTE-
TITLE OF INVENTION: MEDIATED CYTOTOXICITY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/726.607C
FILING DATE: July 10, 1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/460.678
FILING DATE: January 5, 1990
ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2228
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-07-726-607C-1

Query Match 64.7%; Score 19.4; DB 1; Length 2228;
Best Local Similarity 79.3%; Pred. No. 10;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 2 cattatagagctattgctacagataca 30
||||| ||||||| ||||||| ||| |
Db 258 CATCATAGAGCTATTGCTACAGCTGCA 230

RESULT 3
US-07-843-949A-1/c
Sequence 1, Application US/07843949A
Patent No. 5340935
GENERAL INFORMATION:
APPLICANT: Anderson, Paul J.
APPLICANT: Streuli, Michel
APPLICANT: Schlossman, Stuart F.
TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
TITLE OF INVENTION: LYMPHOCYTE-MEDIATED CYTOTOXICITY
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/843.949A
FILING DATE: 19920219
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/460.678
FILING DATE: January 5, 1990
APPLICATION NUMBER: 07/726.607
FILING DATE: July 10, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2228
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-07-843-949A-1

Query Match 64.7%; Score 19.4; DB 1; Length 2228;
Best Local Similarity 79.3%; Pred. No. 10;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 cattatagagcattatgctacagatata 30
||||| ||||| ||||| ||||| |||||
Db 258 CATTCATACAGCTATATGCTGCAGCTGCA 230

RESULT 4
US-08-218-978-1/C
Sequence 1, Application US/08218978
Patent No. 5837811

GENERAL INFORMATION:
APPLICANT: Anderson, Paul J.

APPLICANT: Streuli, Michel

APPLICANT: Schlosman, Stuart F.

TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN

TITLE OF INVENTION: LYMPHOCYTE-MEDIATED CYTOTOXICITY

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB

COMPUTER: IBM PS/2 Model 502 or 555X

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: WordPerfect (Version 5.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/218.978

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/843,949

FILING DATE: February 19, 1992

APPLICATION NUMBER: 07/460,678

FILING DATE: January 5, 1990

APPLICATION NUMBER: 07/726,607

FILING DATE: July 10, 1991

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 00530/021003

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2228

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-218-978-1

Query Match 64.7%; Score 19.4; DB 2; Length 2228;

Best Local Similarity 79.3%; Pred. No. 10;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 cattatagagcattatgctacagatata 30
||||| ||||| ||||| ||||| |||||
Db 258 CATTCATACAGCTATATGCTGCAGCTGCA 230

RESULT 5

US-09-013-881-9

Sequence 9, Application US/09013881

Patent No. 6132964

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Lal, Preeti

APPLICANT: Hillman, Jennifer L.

APPLICANT: Corley, Neil C.

APPLICANT: Guegler, Karl J.

APPLICANT: Shah, Purvi

TITLE OF INVENTION: HUMAN HYDROLASE-LIKE MOLECULES

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/013,881

FILING DATE: HERewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: BILLINGS, LUCY J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0470 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 2162 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: PITUNOT01

CLONE: 094168

US-09-013-881-9

Query Match 63.3%; Score 19; DB 3; Length 2162;
Best Local Similarity 81.5%; Pred. No. 15;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 ttatagagcattatgctacagatata 30
||||| ||||| ||||| ||||| |||||
Db 1815 TTTTGAAGCTATGCGACAGATTCA 1841

RESULT 6

US-08-725-012-1/C

Sequence 1, Application US/08725012

Patent No. 5710025

GENERAL INFORMATION:

APPLICANT: Dikstein, Rivka

APPLICANT: Tjian, Robert

TITLE OF INVENTION: B-Cell Specific Transcription Factor

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

STREET: 268 BUSH STREET, SUITE 3200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,012
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2403
US-08-725-012-1

Query Match 62.7%; Score 18.8; DB 1; Length 2556;
Best Local Similarity 76.7%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 acattatagagctatgtctacagataca 30
||| ||| ||| ||| ||| ||| ||| |||
Db 974 ACAGTTCGTGCTGCTGCACGAACA 945

RESULT 7
US-07-843-949A-3/C
Sequence 3, Application US/07843949A
Patent No. 5340935
GENERAL INFORMATION:
APPLICANT: Anderson, Paul J.
APPLICANT: Streull, Michel
APPLICANT: Schlossman, Stuart F.
TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
TITLE OF INVENTION: LYMPHOCYTE-MEDIATED CYTOTOXICITY
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/843,949A
FILING DATE: 19920219
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/460,678
FILING DATE: January 5, 1990
APPLICATION NUMBER: 07/726,607
FILING DATE: July 10, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1401
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-07-843-949A-3

Query Match 59.3%; Score 17.8; DB 1; Length 1401;
Best Local Similarity 75.9%; Pred. No. 45;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 cattatagagctatgtctacagataca 30
||| ||| ||| ||| ||| ||| ||| |||
Db 253 CATTCATGACAGCTATGACAGAGCTGCA 225

RESULT 8
US-08-218-978-3/C
Sequence 3, Application US/08218978
Patent No. 5837811
GENERAL INFORMATION:
APPLICANT: Anderson, Paul J.
APPLICANT: Streull, Michel
APPLICANT: Schlossman, Stuart F.
TITLE OF INVENTION: DNAS ENCODING PROTEINS ACTIVE IN
TITLE OF INVENTION: LYMPHOCYTE-MEDIATED CYTOTOXICITY
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/218,978
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/843,949
FILING DATE: February 19, 1992
APPLICATION NUMBER: 07/460,678
FILING DATE: January 5, 1990
APPLICATION NUMBER: 07/726,607
FILING DATE: July 10, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/021003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1401
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-218-978-3

RESULT 10
US-08-714-918-42/c
; Sequence 42, Application US/08714918
; Patent No. 6037123
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving

RESULT 11
 US-09-265-315-42/C
 Sequence 42, Application US/09265315
 Patent No. 6187541
 GENERAL INFORMATION:
 APPLICANT: Benton, Bret
 APPLICANT: Lee, Yung J.
 APPLICANT: Malouin, Francois
 APPLICANT: Martin, Patrick K.
 APPLICANT: Schmid, Molly B.
 APPLICANT: Sun, Dongxu
 TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
 TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
 TITLE OF INVENTION: TARGET GENES
 NUMBER OF SEQUENCES: 111
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Lyon & Lyon

STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,315
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Wardburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 2119 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-42

Query Match 58.0%; Score 17.4; DB 4; Length 2119;
Best Local Similarity 77.8%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 ttatagagctatgtctacagataca 30
||||| ||||| ||||| ||||| |||||
Db 446 TTTAAATAGCTTTCTCAGATGATAA 420

RESULT 12
US-09-265-315-42/C
Sequence 42, Application US/09265315
Patent No. 6187541
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ying J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,315
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Wardburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 2119 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-42

Query Match 58.0%; Score 17.4; DB 4; Length 2119;
Best Local Similarity 77.8%; Pred. No. 70;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 4 ttatagagctatgtctacagataca 30
||||| ||||| ||||| ||||| |||||
Db 446 TTTAAATAGCTTTCTCAGATGATAA 420

RESULT 13
US-09-266-417-42/C
Sequence 42, Application US/09266417
Patent No. 6228588
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ying J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: TARGET GENES
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/266,417

Search completed: September 28, 2001, 10:23:38
Job time: 1404 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 2001, 10:00:09 ; Search time 1335.78 Seconds
(without alignments)
212.300 Million cell updates/sec

Title: US-09-551-645-2

Perfect score: 30
Sequence: 1 acattataggagctattgtctacagataca 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est25:*
- 26: gb_est26:*
- 27: gb_est27:*
- 28: gb_est28:*
- 29: gb_est29:*
- 30: gb_est30:*
- 31: gb_est31:*
- 32: gb_est32:*
- 33: gb_est33:*
- 34: gb_est34:*
- 35: gb_est35:*
- 36: gb_est36:*
- 37: gb_est37:*
- 38: gb_est38:*
- 39: gb_est39:*
- 40: gb_est40:*
- 41: gb_est41:*
- 42: gb_est42:*
- 43: gb_est43:*
- 44: gb_est44:*
- 45: gb_est45:*
- 46: gb_est46:*
- 47: gb_est47:*
- 48: gb_est48:*
- 49: gb_est49:*
- 50: gb_est50:*
- 51: gb_est51:*
- 52: gb_est52:*
- 53: gb_est53:*
- 54: gb_est54:*
- 55: gb_est55:*
- 56: gb_est56:*
- 57: gb_est57:*
- 58: gb_est58:*
- 59: gb_est59:*
- 60: gb_est60:*
- 61: gb_est61:*
- 62: gb_est62:*
- 63: gb_est63:*
- 64: gb_est64:*
- 65: gb_est65:*
- 66: gb_est66:*
- 67: gb_est67:*
- 68: gb_est68:*
- 69: gb_est69:*
- 70: gb_est70:*
- 71: gb_est71:*
- 72: gb_est72:*
- 73: gb_est73:*
- 74: gb_est74:*
- 75: gb_est75:*
- 76: gb_est76:*
- 77: gb_est77:*
- 78: gb_est78:*
- 79: gb_est79:*
- 80: gb_est80:*
- 81: gb_est81:*
- 82: gb_est82:*
- 83: gb_est83:*
- 84: gb_est84:*
- 85: gb_est85:*
- 86: gb_est86:*
- 87: gb_est87:*
- 88: gb_est88:*
- 89: gb_est89:*
- 90: gb_est90:*
- 91: gb_est91:*
- 92: gb_est92:*
- 93: gb_est93:*
- 94: gb_est94:*
- 95: gb_est95:*
- 96: gb_est96:*
- 97: gb_est97:*
- 98: gb_est98:*
- 99: gb_est99:*
- 100: gb_est100:*
- 101: gb_est101:*
- 102: gb_est102:*
- 103: gb_est103:*
- 104: gb_est104:*
- 105: gb_est105:*
- 106: gb_est106:*
- 107: gb_est107:*
- 108: gb_est108:*
- 109: gb_est109:*
- 110: gb_est110:*
- 111: gb_est111:*
- 112: gb_est112:*
- 113: gb_est113:*
- 114: gb_est114:*
- 115: gb_est115:*
- 116: gb_est116:*

- 44: em_esthum10:*
- 45: em_esthum11:*
- 46: em_esthum12:*
- 47: em_esthum13:*
- 48: em_esthum14:*
- 49: em_esthum15:*
- 50: em_esthum16:*
- 51: em_esthum17:*
- 52: em_esthum18:*
- 53: em_esthum19:*
- 54: em_esthum20:*
- 55: em_esthum21:*
- 56: em_esthum22:*
- 57: em_esthum23:*
- 58: em_esthum24:*
- 59: em_esthum25:*
- 60: em_esthum26:*
- 61: em_esthum27:*
- 62: em_esthum28:*
- 63: em_esthum29:*
- 64: em_esthum30:*
- 65: em_esthum31:*
- 66: em_esthum32:*
- 67: em_esthum33:*
- 68: em_esthum34:*
- 69: em_esthum35:*
- 70: em_esthum36:*
- 71: em_esthum37:*
- 72: em_esthum38:*
- 73: em_esthum39:*
- 74: em_esthum40:*
- 75: em_esthum41:*
- 76: em_esthum42:*
- 77: em_esthum43:*
- 78: em_esthum44:*
- 79: em_esthum45:*
- 80: em_esthum46:*
- 81: em_esthum47:*
- 82: em_esthum48:*
- 83: em_esthum49:*
- 84: em_esthum50:*
- 85: em_esthum51:*
- 86: em_esthum52:*
- 87: em_esthum53:*
- 88: em_esthum54:*
- 89: em_esthum55:*
- 90: em_esthum56:*
- 91: em_esthum57:*
- 92: em_esthum58:*
- 93: em_esthum59:*
- 94: em_esthum60:*
- 95: em_esthum61:*
- 96: em_esthum62:*
- 97: em_esthum63:*
- 98: em_esthum64:*
- 99: em_esthum65:*
- 100: em_esthum66:*
- 101: em_esthum67:*
- 102: em_esthum68:*
- 103: em_esthum69:*
- 104: em_esthum70:*
- 105: em_esthum71:*
- 106: em_esthum72:*
- 107: em_esthum73:*
- 108: em_esthum74:*
- 109: em_esthum75:*
- 110: em_esthum76:*
- 111: em_esthum77:*
- 112: em_esthum78:*
- 113: em_esthum79:*
- 114: em_esthum80:*
- 115: em_esthum81:*
- 116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		Score		Query		Description	
No.		Score	Match	Length	ID		
1	20.4	68.0	449	229	A04632588	A0462588 HS_5205_2	
2	20.4	68.0	580	250	A2821630	A2821630 ZM0094B07	
3	20.2	67.3	1101	219	CNS006932	AL071602 Drosophila	
4	20	66.7	333	145	BF208312	BF208312 601871444	
5	20	66.7	376	236	A0982143	A0982143 RPTCI-23-*	
6	20	66.7	543	120	AW738482	AW738482 EST339902	
7	20	66.7	589	148	BE394068	BE394068 UP-R-CAO	
8	19.6	65.3	421	115	HS5427	HS5427 Yq96c11.r1	
9	19.6	65.3	444	158	AW3955142	AW3955142 SH4008.r1	
C 10	19.6	65.3	494	220	CNS028PPL	AL144034 Tetradodo	
C 11	19.6	65.3	577	32	AV7092236	AV7092236 AV709226	
12	19.4	64.7	266	26	AV208558	AV208558 AV208558	
13	19.4	64.7	271	27	AV261223	AV261223 AV261223	
14	19.4	64.7	333	250	A2838439	A2838439 ZM0134F07	
15	19.4	64.7	335	14	AA973786	AA973786 oq16d10.s	
C 16	19.4	64.7	409	157	HO6592	HO6592 Y184h01.r1	
17	19.4	64.7	411	234	A0838658	A0838658 HS_5187.E	
18	19.4	64.7	421	16	AI1233300	AI1233300 ov48c11.x	
C 19	19.4	64.7	429	165	BE246597	BE246597 TCBAP1E4	
C 20	19.4	64.7	431	116	AW500680	AW500680 UP-HF-BN0	
C 21	19.4	64.7	432	105	AL0400667	AL0400667 DKFZP434	
C 22	19.4	64.7	440	110	AV748124	AV748124 AV748124	
C 23	19.4	64.7	482	148	BF443814	BF443814 262324.M	
24	19.4	64.7	505	258	TAI18B020	TAI18B020	
25	19.4	64.7	519	235	A0219442	A0219442 T. brucei	
26	19.4	64.7	541	23	A1650696	A0919442 RPTCI-23-	
27	19.4	64.7	583	112	AW139157	A1650696 wB25f04.x	
28	19.4	64.7	633	218	AA550257	AA159157 UP-H-B11	
C 29	19.4	64.7	776	154	BG529306	AA550257 1379m3.g1	
C 30	19.4	64.7	794	234	A0855563	BG529306 6025841.2	
C 31	19.4	64.7	812	144	BE888548	A0855563 CG61655A	
C 32	19.4	64.7	849	144	BF130592	BE888548 601507633	
C 33	19.4	64.7	851	219	CNS013GB	BF130592 60181856	
C 34	19.4	64.7	972	175	BG257075	AL102821 Drosophila	
C 35	19.4	64.7	974	175	BG257151	BG257075 602370633	
C 36	19.4	64.7	1292	166	BE379739	BG257151 602370733	
C 37	19.2	64.0	533	238	AZ081939	BE379739 601159455	
38	19.2	64.0	538	241	AZ288491	AZ081939 UP_340-22	
39	19.2	64.0	560	238	AZ083517	AZ288491 RPTCI-23-	
C 40	19	63.3	170	5	AA319670	AZ083517 RPTCI-23-	
C 41	19	63.3	177	29	AV355923	AA319670 EST19122	
C 42	19	63.3	189	18	AI275684	AV355923 AV355923	
C 43	19	63.3	268	191	Z39625	AI275684 qv66h09.x	
44	19	63.3	277	133	BB423612	Z39625 HSC1P4042.r1	
45	19	63.3	281	160	BB537744	BB423612 BB423612	
						BB537744 BB537744	

ALIGNMENTS

```

JOURNAL      scanning the human genome
MEDLINE      Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
COMMENT      99380589
              Contact: Mahairas GG, Wallace JC, Hood L
              High Throughput Sequencing Center
              University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
              Email: jwallace@u.washington.edu
              Clones are derived from the human BAC library RPCR-11. For BAC
              library availability, please contact Pieter de Jong
              (pieter@econj.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
              or from Resear h Genetics (info@resgen.com). BAC end Web Server:
              http://www.husc.washington.edu
              Plate: 781 row: E column: 22
              Seq primer: SP6
              Class: BAC ends
              High quality sequence stop: 449.

FEATURES
    source
        Location/Qualifiers
            1..449
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate=781 Col=22 Row=E"
                /clone_1bp="RPC1-11 Human Male BAC Library"
                /sex="male"
                /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                Male blood DNA was isolated from one randomly chosen donor
                and partially digested with a combination of EcoRI and
                EcoRI Methylase. Size selected DNA was cloned into the
                pBACe3.6 vector at EcoRI sites"
            BASE COUNT      126 a          70 c          53 g          199 t          1 others
            ORIGIN
                Query Match      68.0%; Score 20.4; DB 229; Length 449;
                Best Local Similarity 80.0%; Pred. No. 98;
                Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
                QY      1 acattatagagctatctgcacagataca 30
                        ||||||| 1 ||||| ||||||| 1
                Db      189 ACATTATTGTGCTTATTCGATACAGATATTA 218

RESULT      2
LOCUS      AZ821630          580 bp      DNA
DEFINITION      2M00094B03R Mouse 10kb plasmid UNGCM1 library Mus musculus genomic
              clone UUGC2M00094B03 R, DNA sequence.
ACCESSION      AZ821630

```

Insert Length: 10000 Std Error: 0.00
 Plate: 0094 row: B column: 03
 Seq primer: CACACAGAAACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 580.

FEATURES
 source
 Location/Qualifiers
 1. 580

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10990"
 /clone="U06C2M034B03"
 /clone_1lb="Mouse 10kb plasmid U06C1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv: Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114(gb|AF129072.1), a copy-number inducible derivative of plasmid p1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
 200 a 115 c 65 g 200 t

ORIGIN

Query Match 68.0%; Score 20.4; DB 250; Length 580;
 Best Local Similarity 80.0%; Pred. No. 1e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 acattatagagctatgctacagataca 30
 ||| ||||| ||||| ||| |||||
 Db 489 ACAATTATAGAGCTATGCTATATATA 460

RESULT 3
 CNS00696 1101 bp DNA GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TERT3 end of BAC:
 DEFINITION BACR32E08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 AL071602
 AL071602.1 GI:4951644
 GSS.
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
 Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of

COMMENT

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

FEATURES
 source
 Location/Qualifiers
 1. 1101

/organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_1lb="RPCI-98"
 /clone="BACR32E08"
 /note="end : TERT3"
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

BASE COUNT 357 a 229 c 236 g 242 t 37 others

ORIGIN

Query Match 67.3%; Score 20.2; DB 219; Length 1101;
 Best Local Similarity 75.9%; Pred. No. 1.3e+02;
 Matches 22; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 cattatagagctatgctacagataca 30
 ||||| ||| ||| ||||| |||
 Db 30 CATTATATGACATATGCTTCAGAMCA 58

RESULT 4
 BF208312 333 bp mRNA EST 06-NOV-2000
 LOCUS 60187144BFI NIH_MGC_53 Homo sapiens CDNA clone IMAGE:4092539 5',
 DEFINITION mRNA sequence.
 ACCESSION BF208312
 VERSION BF208312.1 GI:11101898
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NIH-MGC <http://mgc.ncl.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: image.llnl.gov
 Plate: L1CM949 row: P column: 12
 High quality sequence stop: 226.

FEATURES
 source
 Location/Qualifiers
 1. 333

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4092539"
 /clone_1lb="NIH_MGC_53"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
 site_1: SfiI (ggcgctcgcc); site_2: SfiI (ggcattatggcc)
); Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCGAGCGCGGACATG-dt(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 laboratories (Palo Alto, CA)."

BASE COUNT 136 a 67 c 62 g 68 t
ORIGIN

Query Match 66.7%; Score 20; DB 145; Length 333;
Best Local Similarity 82.1%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 acattataggagctattgctacagata 28
||||| 1 ||||| 1 ||||| 1 ||||| 1
Db 103 ACATTCAGAGAGCTATACCTAAGAAA 130

RESULT 5
AO982143 376 bp DNA GSS 29-JAN-2000
LOCUS AO982143.1
DEFINITION RPI-23-307A9.TV RPI-23 Mus musculus genomic clone RPI-23-307A9,
DNA sequence.
ACCESSION AO982143
VERSION AO982143.1 GI:6814444
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 376)
Zhuo, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akret,
'B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C. M.
Mouse BAC End Sequences from Library RPI-23
Unpublished (1999)
Other-GSS: RPI-23-307A9.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@edjlong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Reseach Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 307 Row: A Column: 9
Seq primer: 77
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..376

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPI-23-307A9"
/clone_11b="RPI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 135 a 53 c 99 g 89 t
ORIGIN

Query Match 66.7%; Score 20; DB 236; Length 376;
Best Local Similarity 82.1%; Pred. No. 1.4e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 2 cattataggagctattgctacagatac 29

Db 107 CATTTATGACGCAATTTCTACTGACAC 80
||||| 1 ||||| 1 ||||| 1 ||||| 1

RESULT 6
AW738482 543 bp mRNA EST 25-APR-2000
LOCUS AW738482
DEFINITION EST1339909 tomato flower buds, anthesis, Cornell University.
Lycopersicon esculentum cDNA clone cTOD7016 5', mRNA sequence.
AW738482
VERSION AW738482.1 GI:7647427
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE 1 (bases 1 to 543)
van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, L.E., Liang
, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Ronning, C.M., Nierman
, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.
Generation of ESTs from tomato flower tissue, anthesis
Unpublished (1999)
Contact: David Fritsch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfritsch@clemson.edu
5 prime sequence.

TITLE JOURNAL
COMMENT
Location/Qualifiers
1..543
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOD7016"
/clone_11b="tomato flower buds, anthesis, Cornell
University"
/tissue_type="flower"
/dev_stage="anthesis"
/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TA496). They
were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."

BASE COUNT 140 a 133 c 113 g 157 t
ORIGIN

Query Match 66.7%; Score 20; DB 120; Length 543;
Best Local Similarity 82.1%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 cattataggagctattgctacagatac 29
||||| 1 ||||| 1 ||||| 1 ||||| 1
Db 158 CATTTATGACCACTGCTCAGAGAC 185

RESULT 7
BF394068 589 bp mRNA EST 27-NOV-2000
LOCUS BF394068
DEFINITION UI-R-CAO-dgy-f-08-0-UI-s1 UI-R-CAO Rattus norvegicus cDNA clone
UI-R-CAO-dgy-f-08-0-UI 3', mRNA sequence.
BF394068
ACCESSION BF394068.1 GI:11378932
VERSION
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

Email: estevenson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Insert length: 949 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 427.

FEATURES

source

1. .444

Location/Qualifiers

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl017-4168"

/clone_11b="Gm-cl017"

/tissue_type="vegetable buds of field grown plants"

/lab_host="X110-Gold"

/note="Vector: pBluescript II Xr; Site_1: EcoRI; Site_2:

XhoI: This cDNA library was constructed from mRNA isolated

from vegetable buds of field grown plants. The cDNA

library was prepared using the Stratagene pBluescript II

Xr library construction kit. Complementary DNA was

synthesized from mRNA using a primer consisting of a poly

(dT) sequence with a XhoI restriction site. EcoRI adapters

were ligated to the blunt-ended cDNA fragments followed by

XhoI digestion. The cDNA fragments were directionally

cloned into the EcoRI-XhoI restriction site of the

pBluescript vector. The ligated cDNA fragments were

transformed into X110-Gold host cells. This library was

constructed by Dr. Randy Shoemaker and Dr. John

Erpelting."

BASE COUNT 88 a 116 c 100 g 138 t 2 others

ORIGIN

Query Match 65.3%; Score 19.6; DB 115; Length 444;

Best Local Similarity 84.6%; Pred. No. 2.2e+02; Mismatches 4; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 427 TTATAGAACCATTCACACAGACAC 402

RESULT 10
 CNS02EPL/c 494 bp DNA GSS 13-MAY-2000
 LOCUS Tetradon nigroviridis genome survey sequence PUC-ori end of clone
 DEFINITION 262112 of library G from Tetradon nigroviridis, genomic survey
 sequence.

ACCESSION AL194034.1 GI:7832140
 VERSION GSS: genome survey sequence.
 KEYWORDS Tetradon nigroviridis.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphae; Acanthopterygii; Percomorphae; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 494)
 Roest-Crolius H., Jallion O., Dasilva C., Fitzames C., Fisher C.,
 Bouneau L., Billault A., Quettier F., Sautin W., Bernot A. and
 Weissenbach J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetradon nigroviridis
 Unpublished
 2 (bases 1 to 494)

REFERENCE
 AUTHORS Roest-Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
 Bernot A., Fitzames C., Winker P., Brotlier P., Quettier F.,
 Sautin W. and Weissenbach J.
 Human gene number estimate provided by genome wide analysis using
 Tetradon nigroviridis DNA sequence
 Unpublished

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

REFERENCE 3 (bases 1 to 494)
 AUTHORS GenomeScope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetradon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetradon.
 Location/Qualifiers

FEATURES

source

1. .494

/organism="Tetradon nigroviridis"

/db_xref="taxon:99883"

/clone="262112"

/clone_11b="G"

/note="Genoscope sequence ID: C0AG262DF06SP1-end ;

BASE COUNT 142 a 104 c 135 g 108 t 5 others

ORIGIN

Query Match 65.3%; Score 19.6; DB 220; Length 494;

Best Local Similarity 78.6%; Pred. No. 2.2e+02; Mismatches 5; Indels 0; Gaps 0;

Matches 22; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db 96 ATTACAGCGCTCATCAGATAGAGA 69

RESULT 11
 AV709226 577 bp mRNA EST 09-OCT-2000
 LOCUS AV709226 ADC Homo sapiens cDNA clone ADCAYC12 5', mRNA sequence.
 DEFINITION AV709226
 ACCESSION AV709226
 VERSION AV709226.1 GI:10726675
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 577)
 Peng Y., Song H., Huang Q., Huang C., Gu Y., Yang Y., Gao G., Xiao
 H., Xu X., Li N., Qian B., Liu F., Qu J., Gao X., Cheng Z., Xu Z.,
 Zeng L., Xu S., Gu W., Tu Y., Jia J., Fu G., Ren S., Zhong M., Lu
 G., Hu R., Chen J., Chen Z., and Han Z.
 Homo sapiens cDNA ADC clones
 Unpublished (2000)
 Contact: Zenguan Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919 (ex.45)
 Fax: 86-21-50801922
 Email: hanzge@hgc.sh.cn
 This clone is available at CHGC in Shanghai.
 Location/Qualifiers

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Oy	4	cttataagactatcgtcacagatca 30
Db	454	TTTTAGACCTATTGNCACAGATTCA 480
RESULT	12	
LOCUS	AV208558	
DEFINITION	AV208558 RIKEN full-length enriched, adult male testis Mus musculus cDNA clone 1700109E04 3', mRNA sequence.	
ACCESSION	AV208558	
VERSION	AV208558.1	GI:6149410
KEYWORDS	EST.	
SOURCE	house mouse. Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; 1 (bases 1 to 266)	
REFERENCE	Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukuishii,Y., Harai,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,D., Kikuchi,N., Koijima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugihara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateo,M., Tomihaga,N., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.	
TITLE	RIKEN Mouse ESTs (Kono,H., et al. 1999)	
JOURNAL	Contact: Yoshihide Hayashizaki	
COMMENT	Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan Tel.: +81-298-36-9013 Fax: +81-298-36-9098 Email: genome-res@rtc.riken.go.jp/ URL:http://genome.rtc.riken.go.jp/ Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998) Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details. Location/Organisms 1. 266 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="1700109E04" /clone_lib="RIKEN full-length enriched, adult male testis" /sex="male" /tissue_type="testis" /dev_stage="adult" /lab_host="SOLR" /note="Site_1: XhoI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was	

BASE COUNT	91 a	33 c	44 g	98 t
ORIGIN				
Query Match	64.7%	Score 19.4;	DB 26;	Length 266;
Best Local Similarity	79.3%	Pred. No. 2.5e+02;		
Matches 23; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0.
QY	1 acattataggaagtcatctgtcaccagataac 29 			
Dd	84 ACATTTTTAAGAGGTATTAACATAAGATTC 112			
RESULT 13				
AV261223	271 bp	mRNA	EST	04-NOV-1999
LOCUS	AY261223 RIKEN full-length enriched, adult male testis (DH10B) Mus			
DEFINITION	musculus CDNA clone 4930419L17 3', mRNA sequence.			
ACCESSION	AY261223			
VERSION	AV261223.1 GI:6248682			
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 271)			
REFERENCE	Kono, H., Alzawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawaji, J., Kikuchi, N., Koijima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Suganohar, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N., Tsunoda, Y., Watanuki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.			
AUTHORS	RIKEN Mouse ESTs (Kono, H., et al. 1999) Unpublished (1999)			
JOURNAL	Contact: Yoshihide Hayashizaki Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan Tel.: +81-298-36-9013 Fax: +81-298-36-9098 Email: genome-res@rtc.riken.go.jp, URL: http://genome.rtc.riken.go.jp/ Sasaki, N., Izawa, M., Watanuki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.			
TITLE	Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998) Itohi, M., Kitsuana, T., Akiyama, J., Shibata, K., Izawa, M., Kawaji, J., Tomari, Y., Carninci, P., Shibata, Y., Shibata, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.			
COMMENT	Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details. Location/Qualifiers 1..271			
FEATURES				
SOURCE				


```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4930419L17"
/clone_1lb="4930419L17"
(DH10B)"
length enriched, adult male testis
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site1: Sali; site2: BamHI. cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAGAGAGATCCAGAGCCCTCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGGAGAGAGATCTCGAGTTAATTAATTAATGCCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pluescript KS(+) after bulk excision from Lambda
FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."

```

Query Match	64.7%	Score 19.4	DB 27	Length 271
Best Local Similarity	79.3%	Pred. No. 2.5e+02		
Matches 23	Conservative 0	Mismatches 6	Indels 0	Gaps 0
QY	1	acattatagagactatctcacagatac	29	
Db	86	ACATTTTAAAGAGGTATACTAAGATFAC	114	
RESULT	14			
AZ838439				
LOCUS				
DEFINITION	AZ838439	323 bp	DNA	
ACCESSION	U01334F07F	Mouse 10kb plasmid	UUGC1M	library
VERSION	clone UUGC2M0134F07	F, DNA sequence.		Mus musculus genomic
KEYWORDS	AZ838439			
ORGANISM	AZ838439.1	GI:13008347		
SOURCE	GSS.			
	house mouse.			
	Mus musculus			
REFERENCE	EMKaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 323)			
	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.			
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid insets			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0134 row: F column: 07 Seq primer: CGTGTAAACGACGCGCCAGT Class: plasmid ends High quality sequence stop: 323. Location/Qualifiers			
FEATURES				

source

1. .323
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG2M0134F07"
/clone_1lb="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g1147321141gblAF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

86 a 68 c 83 g 85 t 1 others

BASE COUNT
ORIGIN

Query Match	64.7%	Score 19.4	DB 250	Length 323
Best Local Similarity	79.3%	Pred. No. 2,66+02		
Matches 23	Conservative 0	Mismatches 6	Indels 0	Gaps 0
Oy	1	acattatagagactatctgcacagatac	29	
Db	232	AGATTTACAGTTGCTATTGCCACACAGATGC	260	
RESULT 15				
AA973786				
LOCUS				
DEFINITION	AA973786	325 bp	mRNA	EST
	oq16d10.1	NCI-CGAP GC4 Homo sapiens cDNA clone IMAGE1586515	3'	07-JUL-1998
ACCESSION	similar to gb:M77142	NUCLEOTYXSIN T1A-1 (HUMAN);		mRNA sequence.
VERSION	AA973786			
KEYWORDS	AA973786.1	GI:3148966		
SOURCE	EST.			
ORGANISM	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 325)			
AUTHORS	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
JOURNAL	Unpublished (1997)			
COMMENT	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-remail.nih.gov			
	Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael			
	Emmert-Buck, M.D., Ph.D.			
	CDNA Library Preparation: M. Bento Soares, Ph.D.			
	CDNA Library Arrayed by: Greg Lennon, Ph.D.			
	DNA Sequencing by: Washington University Genome Sequencing Center			
	Clone distribution: NCI-CGAP clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LIML at:			
	www-bio.llnl.gov/bdrrp/image/image.html			
FEATURES				
	Trace considered overall poor quality			
	Insert Length: 917 Std Error: 0.00			
	Seq Primer: -40m13 fwd. ET from Amersham			
	High quality sequence stop: 1.			
	Location/Qualifiers			

Source

```

1. 325
   /organism:"Homo sapiens"
   /db_xref:"taxon:9606"
   /clone_1:"IMAGE:1586515"
   /clone_1lb:"NCI_CGAP_GC4"
   /tissue_type:"pooled germ cell tumors"
   /lab_host:"DH10B"
   /note:"vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
c119g(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Donaldo."
75 c 60 g 107 t

```

Query Match	Score	DB	Length
64.78;	19.4;	14;	325;
Best Local Similarity	79.38;	Pred No	2 6e+02;

Best Local Similarity 79.3%; Pred. NO. 2.6e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY	2	cattataggaactattgctacagataca	30
Db	239	CATTCAATAGCAGCTAATGCTGCAGCTGCA	267

Search completed: September 28, 2001, 10:22:52
Job time: 1363 sec

Graser
09/551645
Seq IDs 18 2 w/ hite

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2001, 11:18:21 ; Search time 173.48 seconds
(without alignments)
12.788 Million cell updates/sec

Title: US-09-551-645-1
Perfect score: 49
Sequence: 1 TFGATATDTR 10

Scoring table:
BLAST62
Gap 10.0, Gapext 0.5

Searched: 2803329 seqs, 221847457 residues

Total number of hits satisfying chosen parameters: 2803329

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:
1: /cgnl_7/ptodata/1/paa/US06_COMB.pep.*
2: /cgnl_7/ptodata/1/paa/US07_COMB.pep.*
3: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
4: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
5: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
6: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
7: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
8: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
9: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
10: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
11: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
12: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
13: /cgnl_7/ptodata/1/paa/US08_COMB.pep.*
14: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
15: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
16: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
17: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
18: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
19: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
20: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
21: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
22: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*
23: /cgnl_7/ptodata/1/paa/US09_COMB.pep.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	10	US-09-551-645-1	Sequence 1, Appl
2	48	98.0	1612	US-07-742-128-2	Sequence 2, Appl
3	35	71.4	244	US-09-328-352-6861	Sequence 6861, Ap
4	34	69.4	412	US-09-489-039A-9118	Sequence 9118, Ap
5	34	69.4	455	US-09-191-989-4	Sequence 4, Appl
6	34	69.4	482	US-09-438-185-284	Sequence 284, App
7	34	69.4	567	US-09-489-039A-10840	Sequence 10840, A
8	34	69.4	662	US-09-417-507-34454	Sequence 34454, A
9	33	67.3	92	US-09-134-000-5423	Sequence 5423, Ap
10	33	67.3	110	US-09-417-507-28277	Sequence 28277, A

11	33	67.3	173	US-09-489-039A-7555	Sequence 7555, Ap
12	33	67.3	302	US-09-417-507-31001	Sequence 31001, A
13	33	67.3	118	US-09-252-991A-18081	Sequence 18081, A
14	33	67.3	342	US-60-167-217-19670	Sequence 19670, A
15	33	67.3	342	US-60-173-464-16064	Sequence 16064, A
16	33	67.3	342	US-60-191-637-19589	Sequence 19589, A
17	33	67.3	342	US-60-191-681-15455	Sequence 15455, A
18	33	67.3	504	US-09-489-039A-14248	Sequence 14248, A
19	33	67.3	531	US-09-380-420A-2	Sequence 2, Appl
20	33	67.3	531	US-09-380-420B-2	Sequence 2, Appl
21	33	67.3	531	US-09-380-420C-2	Sequence 2, Appl
22	33	67.3	558	US-09-252-691-8968	Sequence 8968, Ap
23	33	67.3	558	US-09-252-691C-8968	Sequence 8968, Ap
24	33	67.3	615	US-60-167-245-745	Sequence 745, App
25	33	67.3	615	US-60-191-637-21026	Sequence 21026, A
26	32	65.3	58	PCT-US01-00663-35001	Sequence 35001, A
27	32	65.3	60	US-60-169-840-5600	Sequence 5600, Ap
28	32	65.3	85	PCT-US01-00663-31565	Sequence 31565, A
29	32	65.3	113	US-60-169-841-1730	Sequence 1730, Ap
30	32	65.3	113	US-60-169-842-3089	Sequence 3089, Ap
31	32	65.3	113	US-60-169-867-4903	Sequence 4903, Ap
32	32	65.3	161	US-60-195-136-1128	Sequence 1128, Ap
33	32	65.3	186	US-60-196-718-7132	Sequence 7132, Ap
34	32	65.3	190	US-60-196-714-700	Sequence 700, App
35	32	65.3	193	US-60-182-093-2408	Sequence 2408, Ap
36	32	65.3	195	US-60-196-713-2230	Sequence 2230, Ap
37	32	65.3	197	US-09-417-507-29323	Sequence 29323, A
38	32	65.3	252	US-09-450-651A-43	Sequence 43, Appl
39	32	65.3	252	US-09-450-651A-44	Sequence 44, Appl
40	32	65.3	252	US-09-450-651A-45	Sequence 45, Appl
41	32	65.3	253	US-09-450-651A-46	Sequence 46, Appl
42	32	65.3	253	US-09-450-651A-47	Sequence 47, Appl
43	32	65.3	271	US-09-252-991A-30514	Sequence 30514, A
44	32	65.3	282	US-09-252-991A-29425	Sequence 29425, A
45	32	65.3	297	PCT-US99-29963-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-09-551-645-1
Sequence 1, Application US/09551645
GENERAL INFORMATION:
APPLICANT: HAHN, Myong-Joon
TITLE OF INVENTION: NEW BIOTOPE TAG RECOGNIZED BY A MONOCLONAL ANTIBODY TO
TITLE OF INVENTION: RICKETTSIA TYPHI
FILE REFERENCE: 105997
CURRENT APPLICATION NUMBER: US/09/551, 645
CURRENT FILING DATE: 2000-04-18
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 10
TYPE: PRT
ORGANISM: Rickettsia typhi
US-09-551-645-1

Handwritten signature

Query Match 100.0%; Score 49; DB 19; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.003; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;
QY 1 TFGATATDTR 10
DB 1 TFGATATDTR 10
RESULT 2
US-07-742-128-2
Sequence 2, Application US/07742128
GENERAL INFORMATION:
APPLICANT: Carl, Mitchell NMI

APPLICANT: Dobson, Michael E.
APPLICANT: Ching, Wei Mei
APPLICANT: Dasch, Gregory A
TITLE OF INVENTION: Gene and Protein Applicable to the
TITLE OF INVENTION: Preparation of Vaccines for and the Detection of
TITLE OF INVENTION: Rickettsia Typhi
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical Res & Dev Cmd
STREET: National Naval Medical Center
City: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/742,128
FILING DATE: 19910809
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: N.C.72,321
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1612 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-742-128-2

Query Match 98.0%; Score 48; DB 3; Length 1612;
Best Local Similarity 90.0%; Pred. No. 1.8;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TFGAATDT 10
||:|||||
Db 1265 TFGAATDT 1274

RESULT 3
US-09-328-352-6861
Sequence 6861, Application US/09328352
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6861
LENGTH: 244
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6861

Query Match 71.4%; Score 35; DB 17; Length 244;
Best Local Similarity 87.5%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FIGAATD 9
||| ||||
Db 172 FIGAATD 179

RESULT 4
US-09-489-039A-9118
Sequence 9118, Application US/09489039A
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709,2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 9118
LENGTH: 412
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9118

Query Match 69.4%; Score 34; DB 18; Length 412;
Best Local Similarity 77.8%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TFGAATD 9
||| ||||
Db 33 TFGAATD 41

RESULT 5
US-09-191-989-4
Sequence 4, Application US/09191989
GENERAL INFORMATION:
APPLICANT: Szasz, Joseph
APPLICANT: Davis, Maria
TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,989
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/738,172
FILING DATE: October 25,1996
APPLICATION NUMBER: 08/465,003
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/575,354
FILING DATE: December 20, 1995
APPLICATION NUMBER: 08/240,158
FILING DATE: May 10, 1994
APPLICATION NUMBER: 08/229,329
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Waidburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/158
TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-191-989-4

Query Match 69.4%; Score 34; DB 15; Length 455;
Best Local Similarity 70.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 TFIGAATDT 10
Db 92 TYNGAATVDT 101

RESULT 6
US-09-438-185-284
Sequence 284, Application US/09438185
GENERAL INFORMATION:
APPLICANT: Stephens, Richard
APPLICANT: Mitchell, Wayne
APPLICANT: Kalman, Sue
APPLICANT: Davis, Ronald
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
FILE REFERENCE: 018941-000411US
CURRENT APPLICATION NUMBER: US/09/438,185
CURRENT FILING DATE: 1999-11-11
PRIOR APPLICATION NUMBER: US 60/108,279
PRIOR FILING DATE: 1998-11-12
PRIOR APPLICATION NUMBER: US 60/128,606
PRIOR FILING DATE: 1999-04-08
NUMBER OF SEQ ID NOS: 1074
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 284
LENGTH: 482
TYPE: PRF
ORGANISM: Chlamydia pneumoniae
US-09-438-185-284

Query Match 69.4%; Score 34; DB 18; Length 482;
Best Local Similarity 88.9%; Pred. No. 3.7e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGAATDT 10
Db 247 FIGAATDT 255

RESULT 7
US-09-489-039A-10840
Sequence 10840, Application US/09489039A
GENERAL INFORMATION:
APPLICANT: Gary Bretton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709,2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 10840
LENGTH: 567
TYPE: PRF

ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10840

Query Match 69.4%; Score 34; DB 18; Length 567;
Best Local Similarity 70.0%; Pred. No. 4.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 TFIGAATDT 10
Db 77 TFIGAATVDT 86

RESULT 8
US-09-417-507-34454
Sequence 34454, Application US/09417507
GENERAL INFORMATION:
APPLICANT: KEITH G. WEINSTOCK ET AL.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
FILE REFERENCE: PAT99-10
CURRENT APPLICATION NUMBER: US/09/417,507
CURRENT FILING DATE: 1999-10-14
NUMBER OF SEQ ID NOS: 44312
SEQ ID NO 34454
LENGTH: 662
TYPE: PRF
ORGANISM: A.fumigatus
US-09-417-507-34454

Query Match 69.4%; Score 34; DB 18; Length 662;
Best Local Similarity 60.0%; Pred. No. 5.4e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 TFIGAATDT 10
Db 140 TYLSAATDT 149

RESULT 9
US-09-134-000-5423
Sequence 5423, Application US/09134000A
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS
FILE REFERENCE: GTC-005
CURRENT APPLICATION NUMBER: US/09/134,000A
CURRENT FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 6810
SEQ ID NO 5423
LENGTH: 92
TYPE: PRF
ORGANISM: Enterococcus faecalis
US-09-134-000-5423

Query Match 67.3%; Score 33; DB 15; Length 92;
Best Local Similarity 87.5%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 IGATATDT 10
Db 40 IGATATDT 47

RESULT 10
US-09-417-507-28277
Sequence 28277, Application US/09417507
GENERAL INFORMATION:
APPLICANT: KEITH G. WEINSTOCK ET AL.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS

```
; TITLE OF INVENTION: FUMIGATUS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 28277
; LENGTH: 110
; TYPE: PRT
; ORGANISM: A.fumigatus
US-09-417-507-28277

Query Match      67.3%; Score 33; DB 18; Length 110;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TFIGAIAT 8
      ||:||||:|
Db      35 TFIGAIST 42

RESULT 11
US-09-489-039A-7555
; Sequence 7555, Application US/09489039A
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7555
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7555

Query Match      67.3%; Score 33; DB 18; Length 173;
Best Local Similarity 87.3%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TFIGAIAT 8
      |||||||
Db      120 TFIGACAT 127

RESULT 12
US-09-417-507-31001
; Sequence 31001, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; TITLE OF INVENTION: FUMIGATUS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 31001
; LENGTH: 212
; TYPE: PRT
; ORGANISM: A.fumigatus
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (57)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-417-507-31001

Query Match      67.3%; Score 33; DB 18; Length 212;
```

```
Best Local Similarity 70.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 TFIGAIATDT 10
      :|||||
Db      79 SFIGAIRTKT 88

RESULT 13
US-09-252-991A-18081
; Sequence 18081, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18081
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18081

Query Match      67.3%; Score 33; DB 16; Length 308;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TFIGAIA 7
      |||||||
Db      154 TFIGAIA 160

RESULT 14
US-60-167-217-19670
; Sequence 19670, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USRS
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19670
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Drosophila
US-60-167-217-19670

Query Match      67.3%; Score 33; DB 23; Length 342;
Best Local Similarity 66.7%; Pred. No. 4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TFIGAIATD 9
      ||:|:|
Db      319 TFIGSEATD 327

RESULT 15
US-60-173-464-16064
; Sequence 16064, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W.D.
```

```

; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16064
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Drosophila
US-60-173-464-16064

```

```

Query Match          67.3%; Score 33; DB 23; Length 342;
Best Local Similarity 66.7%; Pred. No. 4e+02;
-Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 1 TETGAIATD 9
Db 319 TFVGSSEATD 327

```

Search completed: September 28, 2001, 12:30:53
 Job time: 4352 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 28, 2001, 12:25:12 ; Search time 423.26 Seconds
(without alignments)
1.049 Million cell updates/sec

Title: US-09-551-645-1
Perfect score: 49
Sequence: 1 TFIGATATDR 10

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 285759 seqs, 44412722 residues

Total number of hits satisfying chosen parameters: 285759

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA.New:*
1: /cgnl_7/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgnl_7/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgnl_7/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgnl_7/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgnl_7/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgnl_7/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	71.4	419	US-09-573-655A-1647	Sequence 1647, Ap
2	69.4	88	US-09-543-681A-7770	Sequence 7770, Ap
3	69.4	101	PCT-US01-08631-45049	Sequence 45049, A
4	69.4	468	US-09-198-452A-295	Sequence 295, App
5	69.4	548	US-09-543-681A-6920	Sequence 6920, Ap
6	67.3	307	US-09-602-740-390	Sequence 390, App
7	67.3	307	US-09-738-626-5009	Sequence 5009, App
8	67.3	476	PCT-US01-18569-2300	Sequence 2300, Ap
9	66.3	58	US-09-864-761-40366	Sequence 40366, A
10	66.3	85	US-09-864-761-38445	Sequence 38445, A
11	66.3	131	PCT-US01-08631-31597	Sequence 31597, A
12	65.3	131	PCT-US01-08631-31598	Sequence 31598, A
13	65.3	460	US-09-543-681A-8195	Sequence 8195, Ap
14	65.3	475	PCT-US01-08631-60181	Sequence 60181, A
15	65.3	650	PCT-US01-08631-31599	Sequence 31599, A
16	65.3	941	PCT-US01-08631-33592	Sequence 33592, A
17	65.3	949	PCT-US01-08631-53591	Sequence 53591, A
18	65.3	952	PCT-US01-08631-53593	Sequence 53593, A
19	63.3	172	US-09-543-681A-7580	Sequence 7580, Ap
20	63.3	212	US-09-570-581A-1376	Sequence 1376, Ap
21	63.3	320	US-09-738-626-6258	Sequence 6258, Ap
22	63.3	334	US-09-803-110-11209	Sequence 11209, A
23	63.3	412	US-09-902-540-15137	Sequence 15137, A
24	63.3	415	US-60-312-544-5886	Sequence 5886, Ap
25	63.3	430	US-60-312-544-5660	Sequence 5660, Ap
26	63.3	432	US-60-312-544-8168	Sequence 8168, Ap
27	63.3	498	PCT-US01-08631-43542	Sequence 43542, A

28	31	63.3	595	1	PCT-US01-08631-55097	Sequence 55097, A
29	31	63.3	614	5	US-09-543-681A-4330	Sequence 4330, Ap
30	31	63.3	628	5	US-09-803-110-11513	Sequence 11513, A
31	31	63.3	723	1	PCT-US01-14827-9230	Sequence 9230, Ap
32	30.5	62.2	114	5	US-09-834-366-17484	Sequence 17484, A
33	30.5	62.2	152	5	US-09-834-366-17490	Sequence 17490, A
34	30.5	62.2	169	1	PCT-US01-08631-49310	Sequence 49310, A
35	30	61.2	39	5	US-09-864-761-39782	Sequence 39782, A
36	30	61.2	59	5	US-09-764-903-19021	Sequence 19021, A
37	30	61.2	105	5	US-09-757-028-2401	Sequence 2401, Ap
38	30	61.2	108	5	US-09-834-366-18848	Sequence 18848, A
39	30	61.2	122	5	US-09-834-366-13699	Sequence 13699, A
40	30	61.2	176	5	US-09-803-110-11139	Sequence 11139, A
41	30	61.2	296	5	US-09-758-463-1091	Sequence 1091, Ap
42	30	61.2	374	5	US-09-902-540-11888	Sequence 11888, A
43	30	61.2	474	1	PCT-US01-08631-45475	Sequence 45475, A
44	30	61.2	487	1	PCT-US01-08631-47697	Sequence 47697, A
45	30	61.2	500	5	US-09-570-581A-1471	Sequence 1471, Ap

ALIGNMENTS

RESULT 1
US-09-573-655A-1647
Sequence 1647, Application US/09573655A
GENERAL INFORMATION:
APPLICANT: SOLOVEYEV, Victor and TROUKHAN, Maxim
TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptid
TITLE OF INVENTION: Thereby
FILE REFERENCE: 2750-0876P
CURRENT APPLICATION NUMBER: US/09/573, 655A
CURRENT FILING DATE: 2000-05-18
NUMBER OF SEQ ID NOS: 3280
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1647
LENGTH: 419
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-573-655A-1647

Query Match 71.4%: Score 35; DB 5; Length 419;
Best Local Similarity 70.0%: Pred. No. 12;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TFIGATATDR 10
DB 290 SFGSVATDR 299

RESULT 2
US-09-543-681A-7770
Sequence 7770, Application US/09543681A
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543, 681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128, 706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 7770
LENGTH: 88
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-7770

Query Match 69.4%: Score 34; DB 5; Length 88;
Best Local Similarity 87.5%: Pred. No. 3.5;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TFIGAIAT 8
| | | | | | | |
Db 24 TFIGAIAS 31

RESULT 3
PCT-US01-08631-45049
; Sequence 45049, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 45049
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(101)
; OTHER INFORMATION: Xaa ~ X or * as defined in Table 2
PCT-US01-08631-45049

Query Match 69.4%; Score 34; DB 1; Length 101;
Best Local Similarity 70.0%; Pred. No. 4.1;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 TFIGAIATDT 10
| | | | | | | |
Db 46 TWAGAAATDT 55

RESULT 4
US-09-198-452A-295
; Sequence 295, Application US/09198452A
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 295
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-295

Query Match 69.4%; Score 34; DB 5; Length 468;
Best Local Similarity 88.9%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FIGAIATDT 10
| | | | | | | |
Db 233 FIGAIATLT 241

RESULT 5
US-09-543-681A-6920
; Sequence 6920, Application US/09543681A

; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6920
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6920

Query Match 69.4%; Score 34; DB 5; Length 548;
Best Local Similarity 70.0%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 TFIGAIATDT 10
| | | | | | | |
Db 58 TSLGAIATDT 67

RESULT 6
US-09-602-740-390
; Sequence 390, Application US/09602740
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Krieger, Burkhard
; APPLICANT: Schröder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN CARBON METABOLISM AND ENERGY
; FILE REFERENCE: BGI-126CP
; CURRENT APPLICATION NUMBER: US/09/602,740
; CURRENT FILING DATE: 2001-06-20
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 390
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-602-740-390

Query Match 67.3%; Score 33; DB 5; Length 248;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 TFIGAIATD 9
| | | | | | | |
Db 65 TWIGAVGTD 73

RESULT 7
US-09-738-626-5009
; Sequence 5009, Application US/09738626
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO

;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738,626
;; CURRENT FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: Patentin ver. 3.0
;; SEQ ID NO 5009
;; LENGTH: 307
;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5009

Query Match 67.3% Score 33; DB 5; Length 307;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TFGAATDT 9
|||:|
Db 61 TMGAVCTD 69

RESULT 8
PCT-US01-18569-2300
; Sequence 2300, Application PC/TUS0118569
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA133PCT
; CURRENT APPLICATION NUMBER: PCT/US01/18569
; CURRENT FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 2300
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (377)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (437)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (470)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (474)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-18569-2300

Query Match 67.3% Score 33; DB 1; Length 476;
Best Local Similarity 60.0%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TFGAATDT 10
|||:|
Db 450 TFGALAVCT 459

RESULT 9
US-09-864-761-40366
; Sequence 40366, Application US/09864761

;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aecmics-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40366
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008468.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EST HUMAN HIT: A1969295.1, EVALUATE 2.00e-25
; OTHER INFORMATION: SWISSPROT HIT: O95865, EVALUATE 4.60e+00
US-09-864-761-40366

Query Match 65.3% Score 32; DB 5; Length 58;
Best Local Similarity 66.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TFGAATDT 9
|||:|
Db 6 TFGAATDT 14

```
RESULT 10
US-09-864-761-38445
; Sequence 38445, Application US/09864761
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38445
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005609.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: O95865, EVALU8 8.30e+00
; OTHER INFORMATION: EST_HUMAN HIT: A1969295.1, EVALU8 4.00e-41
US-09-864-761-38445
```

Query Match 65.3%; Score 32; DB 5; Length 85;

Best Local Similarity 66.7%; Pred. No. 9; Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TFIGAIATD 9
||:| || |
Db 6 TFIGRIAD 14

```
RESULT 11
PCT-US01-08631-31597
; Sequence 31597, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 31597
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-31597
```

Query Match 65.3%; Score 32; DB 1; Length 131;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TFIGAIATD 9
||:| || |
Db 43 TFIGRIAD 51

```
RESULT 12
PCT-US01-08631-31598
; Sequence 31598, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 31598
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-31598
```

Query Match 65.3%; Score 32; DB 1; Length 131;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TFIGAIATD 9
||:| || |
Db 42 TFIGRIAD 50

RESULT 13
US-09-543-681A-8195
; Sequence 8195, Application US/09543681A

```

; GENERAL INFORMATION:
; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8195
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-8195

Query Match          65.3%; Score 32; DB 5; Length 460;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 FIGAIAAT 8
Db 361 FVGAIAT 367

RESULT 14
PCT-US01-08631-60181
; Sequence 60181, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 60181
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(475)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-08631-60181

Query Match          65.3%; Score 32; DB 1; Length 475;
Best Local Similarity 75.0%; Pred. No. 61;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 IGAIAITD 10
Db 209 IGAVKTDT 216

RESULT 15
PCT-US01-08631-31599
; Sequence 31599, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
```

```

; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 31599
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (178)..(196)
; OTHER INFORMATION: CADHERIN SIGNATURE domain identified by eMATRIX, accession
; OTHER INFORMATION: number PR00205B, p-value=1.706e-11, raw score of 11.39
; NAME/KEY: misc_feature
; LOCATION: (1)...(650)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-08631-31599
```

```

Query Match          65.3%; Score 32; DB 1; Length 650;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TFIGAIAITD 9
Db 21 TFIGRIAD 29
```

Search completed: September 28, 2001, 12:38:05
Job time: 773 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 2001, 10:00:15 ; Search time 1831.18 Seconds
(without alignments)
242.087 Million cell updates/sec

Title: US-09-551-645-2

Perfect score: 30
Sequence: 1 acatttatagagactatgctacagataca 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues

Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents NA.Main:*

1: /cgnl_7/ptodata/1/pna/US06_COMB.seq:*
2: /cgnl_7/ptodata/1/pna/US06_COMB.seq:*
3: /cgnl_7/ptodata/1/pna/US07_COMB.seq:*
4: /cgnl_7/ptodata/1/pna/US08_COMB.seq:*
5: /cgnl_7/ptodata/1/pna/US081_COMB.seq:*
6: /cgnl_7/ptodata/1/pna/US082_COMB.seq:*
7: /cgnl_7/ptodata/1/pna/US083_COMB.seq:*
8: /cgnl_7/ptodata/1/pna/US084_COMB.seq:*
9: /cgnl_7/ptodata/1/pna/US085_COMB.seq:*
10: /cgnl_7/ptodata/1/pna/US086_COMB.seq:*
11: /cgnl_7/ptodata/1/pna/US087_COMB.seq:*
12: /cgnl_7/ptodata/1/pna/US088_COMB.seq:*
13: /cgnl_7/ptodata/1/pna/US089_COMB.seq:*
14: /cgnl_7/ptodata/1/pna/US090_COMB.seq:*
15: /cgnl_7/ptodata/1/pna/US091_COMB.seq:*
16: /cgnl_7/ptodata/1/pna/US092_COMB.seq:*
17: /cgnl_7/ptodata/1/pna/US093_COMB.seq:*
18: /cgnl_7/ptodata/1/pna/US094_COMB.seq:*
19: /cgnl_7/ptodata/1/pna/US095A_COMB.seq:*
20: /cgnl_7/ptodata/1/pna/US095B_COMB.seq:*
21: /cgnl_7/ptodata/1/pna/US095C_COMB.seq:*
22: /cgnl_7/ptodata/1/pna/US095D_COMB.seq:*
23: /cgnl_7/ptodata/1/pna/US096A_COMB.seq:*
24: /cgnl_7/ptodata/1/pna/US096B_COMB.seq:*
25: /cgnl_7/ptodata/1/pna/US096C_COMB.seq:*
26: /cgnl_7/ptodata/1/pna/US096D_COMB.seq:*
27: /cgnl_7/ptodata/1/pna/US096E_COMB.seq:*
28: /cgnl_7/ptodata/1/pna/US097A_COMB.seq:*
29: /cgnl_7/ptodata/1/pna/US097B_COMB.seq:*
30: /cgnl_7/ptodata/1/pna/US097C_COMB.seq:*
31: /cgnl_7/ptodata/1/pna/US098_COMB.seq:*
32: /cgnl_7/ptodata/1/pna/US099_COMB.seq:*
33: /cgnl_7/ptodata/1/pna/US6001_COMB.seq:*
34: /cgnl_7/ptodata/1/pna/US6002_COMB.seq:*
35: /cgnl_7/ptodata/1/pna/US6003_COMB.seq:*
36: /cgnl_7/ptodata/1/pna/US6004_COMB.seq:*
37: /cgnl_7/ptodata/1/pna/US6005_COMB.seq:*
38: /cgnl_7/ptodata/1/pna/US6006_COMB.seq:*
39: /cgnl_7/ptodata/1/pna/US6007_COMB.seq:*
40: /cgnl_7/ptodata/1/pna/US6008_COMB.seq:*
41: /cgnl_7/ptodata/1/pna/US6009_COMB.seq:*
42: /cgnl_7/ptodata/1/pna/US6010_COMB.seq:*
43: /cgnl_7/ptodata/1/pna/US6011_COMB.seq:*

44: /cgnl_7/ptodata/1/pna/US6012_COMB.seq:*
45: /cgnl_7/ptodata/1/pna/US6013_COMB.seq:*
46: /cgnl_7/ptodata/1/pna/US6014_COMB.seq:*
47: /cgnl_7/ptodata/1/pna/US6015_COMB.seq:*
48: /cgnl_7/ptodata/1/pna/US6016_COMB.seq:*
49: /cgnl_7/ptodata/1/pna/US6017_COMB.seq:*
50: /cgnl_7/ptodata/1/pna/US6018_COMB.seq:*
51: /cgnl_7/ptodata/1/pna/US6019_COMB.seq:*
52: /cgnl_7/ptodata/1/pna/US6020_COMB.seq:*
53: /cgnl_7/ptodata/1/pna/US6021_COMB.seq:*
54: /cgnl_7/ptodata/1/pna/US6022_COMB.seq:*
55: /cgnl_7/ptodata/1/pna/US6023_COMB.seq:*
56: /cgnl_7/ptodata/1/pna/US6024_COMB.seq:*
57: /cgnl_7/ptodata/1/pna/US6025_COMB.seq:*
58: /cgnl_7/ptodata/1/pna/US6026_COMB.seq:*
59: /cgnl_7/ptodata/1/pna/US6027_COMB.seq:*
60: /cgnl_7/ptodata/1/pna/US6028_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	30	100.0	30	22 US-09-551-645-2	Sequence 2, Appl1
2	28.4	94.7	5319	3 US-07-742-128-1	Sequence 1, Appl1
3	20.4	68.0	566	49 US-60-177-571-2486	Sequence 2486, Ap
4	20.4	68.0	600	1 PCR-US01-00663-10104	Sequence 10104, A
5	20.4	68.0	6063	52 US-60-205-418-285	Sequence 265, App
6	20	66.7	580	19 US-09-505-532-11257	Sequence 11257, A
7	20	66.7	1217	28 US-09-705-926-25117	Sequence 2517, Ap
8	20	66.7	89138	20 US-09-534-859-149	Sequence 149, App
9	19.6	65.3	434	38 US-60-068-175-649	Sequence 649, App
10	19.6	65.3	930	15 US-09-134-000-3194	Sequence 3194, Ap
11	19.6	65.3	10996	14 US-09-070-921-202	Sequence 202, App
12	19.6	65.3	10996	14 US-09-070-921-202	Sequence 202, App
13	19.6	65.3	31984	36 US-60-045-649-1115	Sequence 1115, App
14	19.6	65.3	31984	36 US-60-045-649-1115	Sequence 1115, App
15	19.6	65.3	32081	38 US-60-068-217-974	Sequence 1106, Ap
16	19.4	64.7	262	17 US-09-304-517A-157812	Sequence 974, App
17	19.4	64.7	262	17 US-09-371-146A-157812	Sequence 157812, App
18	19.4	64.7	266	16 US-09-244-000A-100532	Sequence 100532, App
19	19.4	64.7	268	11 US-08-798-074-10268	Sequence 10268, A
20	19.4	64.7	276	16 US-09-283-466-30204	Sequence 30204, A
21	19.4	64.7	278	23 US-09-616-081-488	Sequence 488, App
22	19.4	64.7	388	17 US-09-362-510A-20613	Sequence 20613, A
23	19.4	64.7	388	17 US-09-362-510A-20613	Sequence 20613, A
24	19.4	64.7	391	28 US-09-710-281-3245	Sequence 3245, App
25	19.4	64.7	406	19 US-09-528-409-1815	Sequence 1815, App
26	19.4	64.7	409	16 US-09-287-618-10305	Sequence 10305, A
27	19.4	64.7	411	22 US-09-572-409-26893	Sequence 26893, A
28	19.4	64.7	430	17 US-09-397-424-1042	Sequence 1042, App
29	19.4	64.7	430	17 US-09-397-424A-1042	Sequence 1042, App
30	19.4	64.7	437	26 US-09-666-355A-13740	Sequence 13740, A
31	19.4	64.7	437	26 US-09-649-164-6809	Sequence 16809, App
32	19.4	64.7	452	15 US-09-287-618-18697	Sequence 18697, A
33	19.4	64.7	467	16 US-09-234-611-7813	Sequence 7813, App
34	19.4	64.7	467	16 US-09-235-076-13128	Sequence 13128, A
35	19.4	64.7	467	16 US-09-248-797-30676	Sequence 30676, A
36	19.4	64.7	467	19 US-09-332-782-13128	Sequence 13128, A
37	19.4	64.7	467	19 US-09-737-223-13128	Sequence 13128, A
38	19.4	64.7	480	16 US-09-234-611-5892	Sequence 5892, App
39	19.4	64.7	480	16 US-09-248-797-42928	Sequence 42928, App
40	19.4	64.7	481	31 US-09-833-790-440	Sequence 440, App
41	19.4	64.7	484	16 US-09-248-797-30675	Sequence 30675, A
42	19.4	64.7	490	25 US-09-644-871-6919	Sequence 6919, App
43	19.4	64.7	501	31 US-09-833-790-4177	Sequence 177, App
44	19.4	64.7	527	18 US-09-474-436-2512	Sequence 2512, App
45	19.4	64.7	536	29 US-09-726-810-860	Sequence 860, App

ALIGNMENTS

```
RESULT 1
US-09-551-645-2
; Sequence 2, Application US/09551645
; GENERAL INFORMATION:
; APPLICANT: HAHN, Myong-Joon
; TITLE OF INVENTION: NEW EPIPTOPE TAG RECOGNIZED BY A MONOCLONAL ANTIBODY TO
; TITLE OF INVENTION: RICKETTSIA TYPHI
; FILE REFERENCE: 105997
; CURRENT APPLICATION NUMBER: US/09/551,645
; CURRENT FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Rickettsia typhi
; FEATURE:
; OTHER INFORMATION: The epitope recognized by SRT10
US-09-551-645-2
```

```
Query Match          100.0%; Score 30; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 acattatagagctattgtctacagataca 30
    ||||||||||||||||||||||||||||
-db 1 acattatagagctattgtctacagataca 30
```

```
RESULT 2
US-07-742-128-1
; Sequence 1, Application US/07742128
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell NMI
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Mei
; APPLICANT: Dasch, Gregory A
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for and the Detection of
; TITLE OF INVENTION: Rickettsia Typhi
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Counsel, Naval Medical Res & Dev Cmd
; STREET: National Naval Medical Center
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/742,128
; FILING DATE: 19910809
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, A. David
; REGISTRATION NUMBER: 24,743
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5319 base pairs
; TYPE: NUCLEIC ACID
```

```
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Rickettsia prowazekii
; STRAIN: Breinl
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 340..345
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 363..368
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 391..5226
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 379..386
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5270..5306
; PUBLICATION INFORMATION:
; AUTHORS: Carl, M.
; AUTHORS: Dobson, M. E.
; AUTHORS: Ching, W.-M.
; AUTHORS: Dasch, G. A.
; TITLE: Characterization of the gene encoding the
; TITLE: protective S-layer protein of Rickettsia
; TITLE: prowazekii; presence of a truncated identical
; TITLE: homolog in rickettsia typhi
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; DATE: 1990
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
US-07-742-128-1
```

```
Query Match          94.7%; Score 28.4; DB 3; Length 5319;
Best Local Similarity 96.7%; Pred. No. 0.076;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 1 acattatagagctattgtctacagataca 30
    ||||||||||||||||||||||||||||
Db 4183 ACATTGTAGAGCTATTGCTACAGATACA 4212
```

```
RESULT 3
US-60-177-571-2486
; Sequence 2486, Application US/60177571
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
; FILE REFERENCE: C1000201
; CURRENT APPLICATION NUMBER: US/60/177,571
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 5082
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2486
; LENGTH: 566
; TYPE: DNA
; ORGANISM: HUMAN
US-60-177-571-2486
```

```
Query Match          68.0%; Score 20.4; DB 49; Length 566;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Oy 1 acattatagagctattgtctacagataca 30
    ||||| || || | ||||| |||||||||
```


DB 238 acattatagagcattgtctacagatata 267

RESULT 4
PCT-US01-00663-10104

; Sequence 10104, Application PC/TUS0100663
; GENERAL INFORMATION:

; APPLICANT: Molecular Dynamics, Inc.

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN PLACENTA

; FILE REFERENCE: PB 0004 NO 7

; CURRENT APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR FILING DATE: 04 February 2000 (04.02.00)

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 26 May 2000 (26.05.00)

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 03 August 2000 (03.08.00)

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 03 October 2000 (03.10.00)

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 27 September 2000 (27.09.00)

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 21 September 2000 (21.09.00)

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 30 June 2000 (30.06.00)

; NUMBER OF SEQ ID NOS: 38837

; SOFTWARE: Molecular Dynamics Sequence Listing Engine

; SEQ ID NO 10104

; LENGTH: 600

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC025119.2

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.52

PCT-US01-00663-10104

Query Match 68.0%; Score 20.4; DB 1; Length 600;

Best Local Similarity 80.0%; Pred. No. 1.6e+02;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 acattatagagcattgtctacagatata 30

DB 361 acattatagagcattgtctacagatata 390

RESULT 5
US-60-205-418-265/c

; Sequence 265, Application US/60205418

; GENERAL INFORMATION:

; APPLICANT: Beasley, Ellen

; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND

; FILE REFERENCE: CLO00056

; CURRENT APPLICATION NUMBER: US/60/205,418

; CURRENT FILING DATE: 2000-05-19

; NUMBER OF SEQ ID NOS: 536

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 265

; LENGTH: 6063

; TYPE: DNA

; ORGANISM: HUMAN

US-60-205-418-265

Query Match 68.0%; Score 20.4; DB 52; Length 6063;

Best Local Similarity 80.0%; Pred. No. 2.4e+02;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 acattatagagcattgtctacagatata 30

DB 3338 ACATTATATTAGAACCATGTTCACAGATACA 3309

RESULT 6
US-09-505-532-11257

; Sequence 11257, Application US/09505532

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Timberlake, William E.

; TITLE OF INVENTION: Plant Genome Sequences and Uses Thereof

; FILE REFERENCE: 38-10(15478)B

; CURRENT APPLICATION NUMBER: US/09/505,532

; CURRENT FILING DATE: 2000-02-16

; PRIOR APPLICATION NUMBER: USSN 60/120,645 1999-02-18; USSN 09/443,025 1999-1

; PRIOR FILING DATE: USSN 60/120,645 1999-02-18; USSN 09/443,025 1999-11-12;

; NUMBER OF SEQ ID NOS: 51470

; SEQ ID NO 11257

; LENGTH: 580

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana columbia

US-09-505-532-11257

Query Match 66.7%; Score 20; DB 19; Length 580;

Best Local Similarity 82.1%; Pred. No. 2.3e+02;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 cattatagagcattgtctacagatata 29

DB 293 ccttatagagcattgtctacagatata 320

RESULT 7
US-09-705-926-2517

; Sequence 2517, Application US/09705926

; GENERAL INFORMATION:

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; TITLE OF INVENTION: ANNOTATED PLANT GENES

; FILE REFERENCE: 38-21(15481)A

; CURRENT APPLICATION NUMBER: US/09/705,926

; CURRENT FILING DATE: 2000-11-06

; NUMBER OF SEQ ID NOS: 21634

; SEQ ID NO 2517

; LENGTH: 1217

; TYPE: DNA

; ORGANISM: Lycopersicon esculentum

US-09-705-926-2517

Query Match 66.7%; Score 20; DB 28; Length 1217;

Best Local Similarity 82.1%; Pred. No. 2.7e+02;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 cattatagagcattgtctacagatata 29

DB 158 cattatagagcattgtctacagatata 185

RESULT 8
US-09-534-859-149

; Sequence 149, Application US/09534859

; GENERAL INFORMATION:

; APPLICANT: Bush, David F.

; APPLICANT: Last, Robert L.

; APPLICANT: Levin, Irena M.

; APPLICANT: Morris, Susan R.

US-09-534-859-149

```

1  APPLICANT: Parnell, Laurence D.
2  APPLICANT: Rounsley, Steven D.
3  APPLICANT: Wiegand, Roger C.
4  TITLE OF INVENTION:      PLANT POLYMORPHIC MARKERS AND USES THEREOF
5  FILE REFERENCE:          38-1015493/B
6  CURRENT APPLICATION NUMBER: US/09/534,859
7  CURRENT FILING DATE:     2000-03-29
8  NUMBER OF SEQ. ID NOS:   1127
9  SEQ ID NO 149
10
11  LENGTH: 89138
12
13  TYPE: DNA
14
15  ORGANISM: Arabidopsis thaliana
16
17  US-09-534-859-149

```

Query Match Similarity	66.7%	Score 20:	DB 20:	length	89138;
Best Local Similarity	82.1%	Pred. No.	5.9e+02		
Matches	23;	Conservative	0;	Mismatches	5;
				Indels	0;
				Gaps	0

Qy 2 cattataggagctattgtctacagatac 29
| ||||| | | | | | |
Db 24610 ccttataggagcattgtcttcagctgc 24637

RESULT 9
 US-60-068-175-649/c
 : Sequence 649, Application us/60068175
 GENERAL INFORMATION:
 APPLICANT: Lagace, Robert E.
 APPLICANT: Corley, Neil C.
 APPLICANT: Russo, Frank D.
 APPLICANT: Hann, Amy L.
 APPLICANT: Finney, Joe D.
 APPLICANT: Gregory L.
 TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF STREPTOCOCCUS PNEUMONIAE
 TITLE OF INVENTION: GENOME, FRAGMENTS THEREOF, AND USES THEREOF
 NUMBER OF SEQUENCES: 1175
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 STREET: 3174 PORTER DRIVE
 CITY: PALO ALTO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: us/60/068,175
 FILING DATE: HEREWITH
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: CERRONE, MICHAEL C.
 REGISTRATION NUMBER: 39,132
 REFERENCE/DOCKET NUMBER: PM-0009-2 P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 855-0555
 TELEFAX: (650) 845-416
 INFORMATION FOR SEQ ID NO: 649:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 434 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: genomic DNA
 IMMEDIATE SOURCE:
 CLONE: SPN2c651
 US-60-068-175-649

Query Match 65.3%; Score 19.6; DB 38; Length 434;

OY	3 attataagagctattgctacagata	28
D0	381 ATGTTACTAGCTTTTGCTACAGATA	356
Matches	22; Conservative	0; Mismatches
		4; Indels
		0; Gaps

```

RESULT 10
US-09-134-000-3194/c
; Sequence 3194, Application US/09134000A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS
; TITLE OF INVENTION: FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-005
; CURRENT APPLICATION NUMBER: US/09/134, 000A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 6810
; SEQ ID NO 3194
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000-3194

```

Query Match	65.3%	Score 19.6;	DB 15;	Length 930;
Best Local Similarity	84.6%	Pred. No. 3.8e+02;		
Matches 22; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

```

Oy      3 attatagagcattgtctacagata 28
          || ||||| ||| ||||| ||||
Db      466 ATGTATAGTAGCTTTTGCTAGAGATA 441

```

```

RESULT 11
US-09-070-927-202/c
: Sequence 202, Application US/09070927
: GENERAL INFORMATION:
: APPLICANT: Charles Kunsch
: APPLICANT: Patrick J. Dillon
: APPLICANT: Steven C. Barash
: TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and
: TITLE OF INVENTION: Polypeptides
: NUMBER OF SEQUENCES: 982
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Disquette, 3.50 Inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/070,927
: FILING DATE: herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: A. Anders Brookes
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB369
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 202:
: SEQUENCE CHARACTERISTICS:

```



```

: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/60/046,653
: FILING DATE: HEREWITH
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: CERRONE, MICHAEL C.
: REGISTRATION NUMBER: 39,132
: REFERENCE/DOCKET NUMBER: PM-0003-1 P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 855-0555
: TELEFAX: (415) 845-4166
: INFORMATION FOR SEQ ID NO: 1106:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 31984 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: IMMEDIATE SOURCE:
: CLONE: EFALc1106
: US-60-046-653-1106

```

```

Query Match 65.3%; Score 19.6; DB 36; Length 31984;
Best Local Similarity 84.6%; Pred. No. 7.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 3 attataggagctattgtctacagata 28
   ||||||| ||||| ||||| |||||
Db 8280 ATGTATAGTACTTTTGTCTAGAGATA 8255

```

```

RESULT 15
US-60-068-217-974/C
: Sequence 974, Application US/60068217
: GENERAL INFORMATION:
: APPLICANT: Lagace, Robert E.
: APPLICANT: Corley, Neil C.
: APPLICANT: Russo, Frank D.
: APPLICANT: Hann, Amy L.
: APPLICANT: Heath, Joe D.
: APPLICANT: Finney, Gregory L.
: TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF ENTEROCOCCUS FAECALIS
: NUMBER OF SEQUENCES: 1239
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 PORTER DRIVE
: CITY: PALO ALTO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/60/068,217
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: CERRONE, MICHAEL C.
: REGISTRATION NUMBER: 39,132
: REFERENCE/DOCKET NUMBER: PM-0003-5 P
: TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: (650) 855-0555
: TELEFAX: (650) 845-4166
: INFORMATION FOR SEQ ID NO: 974:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 32081 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: genomic DNA
: IMMEDIATE SOURCE:
: CLONE: EFALc974
: US-60-068-217-974

```

```

Query Match 65.3%; Score 19.6; DB 38; Length 32081;
Best Local Similarity 84.6%; Pred. No. 7.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 3 attataggagctattgtctacagata 28
   ||||||| ||||| ||||| |||||
Db 8240 ATGTATAGTACTTTTGTCTAGAGATA 8215

```

```

Search completed: September 28, 2001, 10:56:19
Job time: 3364 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 2001, 10:23:37 ; Search time 4038.06 Seconds
(without alignments)
15.005 Million cell updates/sec

Title: US-09-551-645-2

Perfect score: 30
1 acattataggagctattgctacagataca 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1770485 seqs, 1009842910 residues

Total number of hits satisfying chosen parameters: 3540970

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgnl_7/ptodata/2/pna/US06_NEW_COMB.seq:*
2: /cgnl_7/ptodata/2/pna/US07_NEW_COMB.seq:*
3: /cgnl_7/ptodata/2/pna/US08_NEW_COMB.seq:*
4: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq:*
5: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq:*
7: /cgnl_7/ptodata/2/pna/US09_NEW_COMB.seq:*
8: /cgnl_7/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.4	68.0	600	7	US-09-864-761-12782
2	20	66.7	89138	6	US-09-803-736-149
3	19.6	65.3	630	8	US-60-253-652-7674
4	19.4	64.7	268	4	US-08-798-074B-10268
5	19.4	64.7	268	4	US-08-798-074B-10268
6	19.4	64.7	268	4	US-08-798-074C-10268
7	19.4	64.7	395	8	US-60-253-378-12293
8	19.4	64.7	467	6	US-09-904-809-7813
9	19.4	64.7	480	6	US-09-904-809-5892
10	19.4	64.7	500	6	US-09-652-125A-2073
11	19.4	64.7	778	6	US-09-652-125A-8012
12	19.4	64.7	2253	1	PCR-US01-08631-16714
13	19.4	64.7	4873	7	US-09-838-601-3927
14	19.4	63.3	445	7	US-09-880-107-3525
15	19.4	63.3	461	6	US-09-904-703-6861
16	19.4	63.3	789	7	US-09-764-905-36889
17	19.4	63.3	810	7	US-09-764-905-7324
18	19.4	63.3	2006	6	US-09-898-888-907
19	19.4	63.3	2311	1	PCR-US01-08631-21439
20	19.4	63.3	86121	6	US-09-803-736-777
21	19.4	63.3	114418	6	US-09-803-736-234
22	18.8	62.7	306	6	US-09-758-466-264
23	18.8	62.7	306	6	US-09-813-154-748
24	18.8	62.7	306	7	US-09-912-292-39947
25	18.8	62.7	307	7	US-09-317-311C-234

C 26	18.8	62.7	390	5	US-09-824-559-5112	Sequence 5112, Ap
C 27	18.8	62.7	499	7	US-09-846-040-335	Sequence 335, App
C 28	18.8	62.7	594	7	US-09-824-559-8537	Sequence 8537, Ap
C 29	18.8	62.7	622	7	US-09-846-040-344	Sequence 344, App
C 30	18.8	62.7	2288	6	US-09-620-312B-24	Sequence 24, Appl
C 31	18.8	62.7	2549	6	US-09-764-898-82	Sequence 82, Appl
C 32	18.8	62.7	3468	1	PCR-US01-08631-9459	Sequence 9459, Ap
C 33	18.8	62.7	4772	8	US-60-278-232-4648	Sequence 4648, Ap
C 34	18.8	62.7	5484	6	US-09-760-466-1631	Sequence 1631, Ap
C 35	18.8	62.7	6868	6	US-09-808-383-15569	Sequence 5569, Ap
C 36	18.8	62.7	29334	6	US-09-760-466-1632	Sequence 1632, Ap
C 37	18.8	62.7	102299	6	US-09-803-736-1216	Sequence 1216, Ap
C 38	18.8	62.7	105863	6	US-09-803-736-1145	Sequence 1145, Ap
C 39	18.6	62.0	328	8	US-60-253-457-13288	Sequence 13288, A
C 40	18.6	62.0	553	8	US-60-253-652-10317	Sequence 10317, A
C 41	18.6	62.0	1985	8	US-60-278-561-15517	Sequence 15517, A
C 42	18.6	62.0	110714	6	US-09-803-736-1467	Sequence 1467, Ap
C 43	18.6	62.0	110714	6	US-09-803-736-1467	Sequence 1467, Ap
C 44	18.4	61.3	374	4	US-08-798-074B-12588	Sequence 12588, A
C 45	18.4	61.3	374	4	US-08-798-074B-12588	Sequence 12588, A

ALIGNMENTS

RESULT 1
US-09-864-761-12782
; Sequence 12782, Application US/09864761
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180, 312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207, 456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632, 366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236, 359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234, 687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608, 408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774, 203

```

; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12782
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC025119.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.45
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.54
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.56
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
US-09-864-761-12782

Query Match          68.0%; Score 20.4; DB 7; Length 600;
Best Local Similarity 80.0%; Pred. No. 14;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy      1 acattataggagctatgtctacagatata 30
Db      361 acattaatagaaccatgtctacagatata 390

RESULT 2
US-09-803-736-149
; Sequence 149, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(13493)D
; CURRENT APPLICATION NUMBER: US/09/803,736
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/534,859
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 04983.0206CPUS01 38-10
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 149
; LENGTH: 89138
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-803-736-149

Query Match          66.7%; Score 20; DB 6; Length 89138;
Best Local Similarity 82.1%; Pred. No. 50;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy      2 cattataggagctatgtctacagatata 29
Db      24610 ccttataggagcgtatgtctacagctgc 24637

RESULT 3
US-60-253-652-7674/c
; Sequence 7674, Application US/60253652
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R
; TITLE OF INVENTION: Compositions isolated from bovine
; FILE REFERENCE: 1055P2
; CURRENT APPLICATION NUMBER: US/60/253,652
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 29954
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 7674
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Bovine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(630)
; OTHER INFORMATION: n = A,T,C or G
US-60-253-652-7674

Query Match          65.3%; Score 19.6; DB 8; Length 630;
Best Local Similarity 81.5%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy      4 ttataggagctatgtctacagatata 30
Db      627 TGTATAGCAGCATTTCTACAAATANA 601

RESULT 4
US-08-798-074B-10268/c
; Sequence 10268, Application US/08798074B
; GENERAL INFORMATION:
; APPLICANT: Rosen, Craig A.
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Li, Haodong
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 22
; FILE REFERENCE: PO-22
; CURRENT APPLICATION NUMBER: US/08/798,074B
; PRIOR FILING DATE: 1997-02-12
; PRIOR APPLICATION NUMBER: 60/011,617
; PRIOR FILING DATE: 1996-02-13
; NUMBER OF SEQ ID NOS: 13449
; SEQ ID NO 10268
; LENGTH: 268
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (118)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (206)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (220)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (226)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (235)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (264)
; OTHER INFORMATION: n equals a,t,g, or c
US-08-798-074B-10268

Query Match          64.7%; Score 19.4; DB 4; Length 268;
Best Local Similarity 79.3%; Pred. No. 33;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy      2 cattataggagctatgtctacagatata 30
Db      49 CATTTATAGCATTTATGCTATGATGATAA 21

RESULT 5
US-08-798-074B-10268/c
; Sequence 10268, Application US/08798074B
```



```

; GENERAL INFORMATION:
; APPLICANT: Rosen, Craig A.
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Li, Haodong
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 22
; FILE REFERENCE: PO-22
; CURRENT APPLICATION NUMBER: US/08/798,074B
; CURRENT FILING DATE: 1997-02-12
; PRIOR APPLICATION NUMBER: 60/011,617
; PRIOR FILING DATE: 1996-02-13
; NUMBER OF SEQ ID NOS: 13449
; SEQ ID NO 10268
; LENGTH: 268
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (118)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (206)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (220)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (226)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (235)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (264)
; OTHER INFORMATION: n equals a,t,g, or c
; US-08-798-074B-10268

Query Match          64.7%; Score 19.4; DB 4; Length 268;
Best Local Similarity 79.3%; Pred. No. 33;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 cattatagagctattgtcacagataca 30
DB 49 CATTATAGCATTATTGCTAATGATMAA 21

RESULT 6
US-08-798-074C-10268/C
; Sequence 10268, Application US/08798074C
; GENERAL INFORMATION:
; APPLICANT: Rosen, Craig A.
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Li, Haodong
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 22
; FILE REFERENCE: PO-22
; CURRENT APPLICATION NUMBER: US/08/798,074C
; CURRENT FILING DATE: 1997-02-12
; PRIOR APPLICATION NUMBER: 60/011,617
; PRIOR FILING DATE: 1996-02-13
; NUMBER OF SEQ ID NOS: 13449
; SEQ ID NO 10268
; LENGTH: 268
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (118)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (206)
; OTHER INFORMATION: n equals a,t,g, or c

```

```

; NAME/KEY: misc_feature
; LOCATION: (220)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (226)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (235)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (264)
; OTHER INFORMATION: n equals a,t,g, or c
; US-08-798-074C-10268

Query Match          64.7%; Score 19.4; DB 4; Length 268;
Best Local Similarity 79.3%; Pred. No. 33;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 cattatagagctattgtcacagataca 30
DB 49 CATTATAGCATTATTGCTAATGATMAA 21

RESULT 7
US-60-253-378-12293
; Sequence 12293, Application US/60253378
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J
; TITLE OF INVENTION: Polynucleotides isolated from plants
; FILE REFERENCE: 1054P3
; CURRENT APPLICATION NUMBER: US/60/253,378
; CURRENT FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 40367
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12293
; LENGTH: 395
; TYPE: DNA
; ORGANISM: Pinus radiata
; US-60-253-378-12293

Query Match          64.7%; Score 19.4; DB 8; Length 395;
Best Local Similarity 79.3%; Pred. No. 36;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 cattatagagctattgtcacagataca 30
DB 313 ccttgatgagactatgtcgcataatata 341

RESULT 8
US-09-904-809-7813/C
; Sequence 7813, Application US/09904809
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-757
; CURRENT APPLICATION NUMBER: US/09/904,809
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 09/234,611
; PRIOR FILING DATE: 1999-01-22
; NUMBER OF SEQ ID NOS: 21025
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7813
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(467)

```


Db 258 CATTCATAGCAGCTAATGCTGCGAGCTGCA 230

RESULT 13

US-09-838-601-3927/c
; Sequence 3927, Application US/09838601

; GENERAL INFORMATION:

; APPLICANT: Gearling, David P.

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM A

; FILE REFERENCE: 1600.1042-002
; TITLE OF INVENTION: MIDTERM PLACENTA LIBRARY

; CURRENT FILING DATE: 2001-04-19

; PRIOR APPLICATION NUMBER: US/09/838,601

; PRIOR FILING DATE: 1998-09-18

; PRIOR APPLICATION NUMBER: US 60/100,944

; PRIOR FILING DATE: 1998-11-05

; PRIOR APPLICATION NUMBER: US 60/126,902

; PRIOR FILING DATE: 1999-03-30

; PRIOR APPLICATION NUMBER: US 09/397,761

; NUMBER OF SEQ ID NOS: 4410

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3927

; LENGTH: 4873

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(4873)

; OTHER INFORMATION: n = A,T,C or G

*US-09-838-601-3927

Query Match 64.7%; Score 19.4; DB 7; Length 4873;

Best Local Similarity 79.3%; Pred. No. 56;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 cattataggagctattgctacagatata 30

Db 443 CATTCATAGCAGCTAATGCTGCGAGCTGCA 415

RESULT 14

US-09-880-107-3525/c

; Sequence 3525, Application US/09880107

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherf, Uwe

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880,107

; PRIOR FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: US 60/237,054

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3525

; LENGTH: 445

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(445)

; OTHER INFORMATION: n = A,T,C or G

*US-09-880-107-3525

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 ttataggagctattgctacagatata 30

Db 364 TTTTAAAGCTAATGGGACAGATTC A 338

RESULT 15

US-09-904-703-6861/c

; Sequence 6861, Application US/09904703

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; FILE REFERENCE: 20411-758CON1

; CURRENT APPLICATION NUMBER: US/09/904,703

; PRIOR FILING DATE: 2001-07-12

; PRIOR APPLICATION NUMBER: 09/210,298

; PRIOR FILING DATE: 1998-12-09

; PRIOR APPLICATION NUMBER: US 09/17812

; NUMBER OF SEQ ID NOS: 17812

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 6861

; LENGTH: 461

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(461)

; OTHER INFORMATION: n = A,T,C or G

*US-09-904-703-6861

Query Match 63.3%; Score 19; DB 6; Length 461;

Best Local Similarity 81.5%; Pred. No. 56;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 ttataggagctattgctacagatata 30

Db 49 TTGATAGCCGCTTTTCTCTACAGATAGA 23

Search completed: September 28, 2001, 12:25:03
Job time: 7286 sec

